Theoretical modelling of sialoside hydrolysis: reaction intermediates and mechanistic aspects

Firth-Clark, Stuart

Award date:
1999

Awarding institution:
University of Bath

THEORETICAL MODELLING OF SIALOSIDE HYDROLYSIS:

REACTION INTERMEDIATES AND MECHANISTIC ASPECTS

Submitted by Stuart Firth-Clark
for the degree of PhD
of the University of Bath
1999

"We have different gifts, according to the grace given us. If a person's gift is prophesying, let them use it in proportion to their faith. If it is serving, let them serve; if it is teaching let them teach; if it is encouraging, let them encourage; if it is contributing to the needs of others, let them give generously; if it is leadership, let them govern diligently; if it is showing mercy, let them do it cheerfully."

Romans 12:6-8

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

THEORETICAL MODELLING OF SIALOSIDE HYDROLYSIS:
Reaction Intermediates and Mechanistic Aspects

CONTENTS

ABSTRACT

This thesis considers two aspects of the reaction involving the hydrolysis of sialosides using theoretical modelling techniques. Firstly, the structure and stability of possible reaction intermediates in the non-enzymic reaction and secondly, an investigation into the mechanism of sialidase action.

Ab initio levels of theory are used to investigate the structure and stability of models representing the three possible intermediates which result on the departure of the leaving group from the sialic acid moiety. The structure and stability of the hydroxyl-substituted α-lactone, hydroxyoxiranone, is considered in the gas phase and is placed in the context of known α-lactone chemistry. The relative stability of the charged-separated zwitterion, hydroxyoxiranone and hydroxycarbene/CO_2 systems are considered *in vacuo* and in a continuum environment. A larger model that incorporates the pyranose ring is also used.

Explicit solvation of the full sialyl α-lactone and zwitterion intermediates is described using the AM1/TIP3P hybrid potential. A technique is devised to consider ten different solvent environments for both intermediates in which five solvent environments are initially generated around each intermediate via simulated annealing. The related α-lactone and zwitterion species are generated from the connecting TS structure for the α-lactone ring opening/closure within each environment. This allows investigation into the reproducibility of modelling chemical changes using the hybrid QM/MM method. This work also provides the energy change between both intermediates within the explicit solvent environment.

The QM/MM potential is also used to model the sialidase action on sialosides. Two steps are specifically considered: the proton transfer from Asp(H)-151 to the glycoside oxygen on the substrate, and the proton abstraction by Asp-151 following addition of water to the bound zwitterion intermediate. The transition structures are given for both processes and calculated KIEs are compared with experimental values.

ACKNOWLEDGEMENTS

I wish to acknowledge the following people who have helped me over the past few years.

Firstly, to Ian and Jan Williams for their personal and professional support. To Ian for his top-class supervision throughout the Ph.D., and for his informal "chats" that have been both, informative and supportive. To Jan for the countless trips to Sainsbury's which kept the kitchen well stocked, thank you.

To Dr. Vicent Moliner and Dr. Alex Turner for their helpful discussions on QM/MM modelling and all the personal support that they have given. To Dr. Christopher Rodriquez for his helpful discussions on high level *ab initio* modelling as well as showing me what the city of Bath has to offer. To Gus Ruggiero for keeping me company over these past 3-years or so. We got there in the end! To Gail Rickard and Stuart Gooding for their personal and professional support; to Cristian Lapthorn and Anders Hogner for their interest in the sialidase work. *I thank you all for such memorable and enjoyable times.*

I also thank the Computational Chemistry Working Party (UK) for granting me time allocations on the DEC 8400 at Rutherford-Appleton Laboratory (R.A.L.), Oxford, UK. The majority of the work in this thesis was done using these allocations, without which the calculations presented here would have been computationally intractable. I also thank the Human Resources department at Bridge Information Systems for allowing me to use their printing facilities.

Finally, I would like to thank the EPSRC for financial support.

I would like to dedicate this thesis to MUM and DAD.

1. PROLOGUE

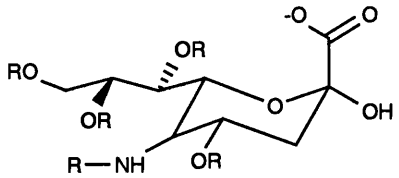
Anyone involved with the subject of "chemistry" will know how exhilarating it is. From the investigations into chemical structure to the synthesis of new compounds, chemistry provides an environment in which people can enjoy and partake in a diverse range of scientific investigations.

Perhaps this is because chemistry works in conjunction with other sciences such as physics and biology. In some ways, it can be considered as the application of the theories of physics (such as quantum mechanics), while the subject also underpins many of the issues associated with biology. Understanding the process of viral infections at the molecular level, for example, can result in the generation of therapeutics for certain diseases. Therapeutics are often designed with prior knowledge of molecular structure (involving X-ray crystallographers and molecular modellers) and are usually made via a complex set of chemical reactions (by organic and medicinal chemists). Physical chemists provide understanding on how the energy of a chemical system changes throughout the course of a reaction. This in turn gives information about the speed of a chemical reaction and the stability of various chemical species associated with such a change.

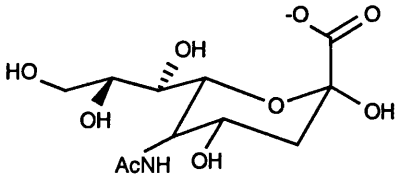
In chemistry, the central aim is to understand the changes taking place in a chemical system, especially at the structural level. Chemists often ask the questions, "what is the molecular structure of the system I am dealing with?", "why is this structure more stable than that structure", "how does the molecular structure alter throughout the course of a reaction?", "how does the energy of the system vary as these changes take place?" and "how fast will this reaction go?". Despite the advancement in technology, however, many of these questions can not be answered using practical chemistry alone because many chemical structures of interest are elusive to current techniques.

Theoretical chemistry can use established theories of physics to model chemical structure and reactivity. The investigations undertaken in this thesis concentrate on modelling one type of reaction: the attack of water on a series of biologically

important molecules known as sialosides, or sialic acids, **1**. The simplest derivative of these is N-acetyl-α-D-neuraminic acid, **2**, and is the species associated with the work presented herein. It is commonly found as a terminal residue on cell surfaces and is therefore a receptor for many biological processes (such as influenza virus infection of a host organism). This thesis thus lies on the intersection of physics, chemistry and biology.



1



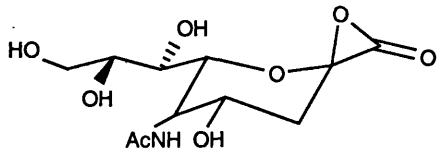
2

However, theories only provide a model to explain the facts; the facts are not generated to explain the theory! Increasingly complex levels of theory will tend to describe facts more accurately since they make fewer approximations. The complexity of theory used for a simulation will often depend on the size of the molecular system under study and the amount of computer power available to carry out the simulation. Because approximations have to be made when describing a chemical system, theoretical results need to be compared with experimentally determined properties (wherever possible) to ensure the simulation is performing correctly. When used with care, theory is able to provide further insight into molecular species and their associated reactions. This is one of the major advantages of modelling the hydrolysis of the sialoside system: there exists experimental data to compare with calculated theoretical results.[1,2]

Theory is unfortunately riddled with jargon. Thus, explanation of the basic concepts used to generate the results reported within these pages is first discussed in *Generation, Exploration and Analysis of Potential Energy Surfaces* (chapter 2). Armed with the terminology, an outline as to how the theoretical simulations were set up is discussed in *Simulation Methods and Approaches* (chapter 3). Specific information on the set up of particular simulations is given when that particular piece of work is being reported.

The main body of the thesis could be considered as two distinct sections, each investigating a different aspect of sialoside hydrolysis. The first section considers the "structure and stability of intermediates associated with the non-enzymic reaction" using various levels of complexity. The second part of the thesis considers "the reaction mechanism associated with the enzyme catalysed hydrolysis of sialosides." Every results chapter contains an introduction to the work under investigation and places it in context with the current scientific knowledge.

One of the issues arising from the experimental investigations into the non-enzymic hydrolysis reaction is the possible formation of an α-lactone intermediate, **3**.[1] Such three-membered ring cyclic esters are interesting because they have a high degree of ring strain and are therefore unstable species. What effects stabilise the sialyl α-lactone in order for the reaction to proceed through such an unstable intermediate? To answer this question, a much smaller system is used consisting of only the basic α-lactone ring and a hydroxyl substituent. Using a smaller system allows the use of much more complex levels of theory. In the *Structure and Stability of Hydroxyoxiranone* (chapter 4), an investigation is undertaken into the chemical structure of the hydroxyl substituted α-lactone and predicts the heat of formation for the molecule using high levels of theory. Therefore, "how stable is this molecule?" and "what effect does the hydroxyl group have on the ring strain?" are the two main questions considered in this chapter. An interesting side issue considered in this work is how various other α-substituents may affect the ring strain of the system.



3

Conceivably, two other intermediates could also exist in the non-enzymic reaction: a zwitterion (which results from the opening of the three-membered α-lactone ring) and a carbene which results from the decarboxylation of the zwitterion. The *Intermediates in Non-enzymic Sialoside Hydrolysis* (chapter 5) takes a close look at the three possible intermediates using models related to the hydroxyoxiranone structure. The major question asked here is "which one of these structures is the most stable?" and "how important is the pyranose ring oxygen in stabilising the three

intermediates?" Although it is interesting to consider these issues about relative stability in the gas phase, it should be remembered that the sialoside hydrolysis reaction takes place in an aqueous environment. Therefore, chapter 5 also considers the effect of aqueous solvent on the relative stability of the three intermediates: "how does solvent alter the relative stability of the three intermediates?" These issues are also considered with a larger model that also incorporates the whole pyranose ring; this provides a check to see if the calculated results are "model-dependent".

The problem in using these simpler models is the risk of neglecting the importance of the remaining system. However, as simulations become more "chemically realistic" a less complex level of theory has to be used to alleviate "computational cost". This is certainly the case in the *Explicit Solvation of the Sialyl α-Lactone and Zwitterion Intermediates* (chapter 6) which considers the interconversion of the two intermediates. In this work, the full sialyl system is used to describe the zwitterion and α-lactone intermediates within an aqueous environment described by *ca.* 500 water molecules. A semi-empirical quantum mechanical method is used to describe the solute while a much simpler classical model is used to model the solvent. Two major questions are tackled here: "does the relative stability of these two intermediates alter by using a more realistic chemical system described in a more approximate way?" A further complication of using a more "chemically realistic" system is the increased flexibility associated with it. This chapter, therefore, pertinently asks, "can reproducible results be obtained using a QM/MM potential despite the very flexible aqueous environment?" To consider this issue, a simulation method is devised to model the α-lactone/zwitterion interconversion within different solvent environments.

The second section of the thesis begins with a discussion on how enzyme catalysed hydrolysis of sialosides is important to influenza infection. *Structure, Function and Inhibition of Influenza Sialidase* (chapter 7) gives some background into the enzyme's structure, its role in the virus' infectious cycle, and the possible ways of hindering its action. Research into the latter area is currently intensive, as specific sialidase inhibitors would have the ability to delay the onset of the disease. An effective inhibitor would be one that hinders the enzyme from catalysing the

sialoside hydrolysis reaction. However, the mechanism by which this reaction takes place at molecular level is still largely unknown. Molecular modelling is currently in a position to provide some further insight into such processes.

Therefore, in *QM/MM Modelling of Influenza-A Sialidase Action on Sialosides* (chapter 8) a QM/MM method is used to investigate the sialoside hydrolysis mechanism under enzymic conditions. Again, a semi-empirical QM method is used to model the reaction centre while the remaining part of the system is described classically. The chapter brings together previous data and postulations of the mechanism and constructs a possible mechanism in which certain steps are investigated.

As hinted already, this thesis constantly asks questions about the success and reliability of the models being used. Certainly, these questions dominate the discussion and results given in chapters 6 and 8 on the QM/MM methodology. Developing these simulations further is discussed in *Conclusions and Possibilities for Further Study* (Chapter 9). This part of the thesis provides a "general conclusion"; specific conclusions about the work considered here are given in the relevant chapters. More importantly, this chapter suggests possible approaches for future simulations.

The last part of the thesis lists the *references* used throughout while the *appendices* give examples of the program scripts used to generate the data and information reported herein.

2. GENERATION, EXPLORATION AND ANALYSIS OF POTENTIAL ENERGY SURFACES

2.1. Potential Energy Surfaces[3,4]

A potential energy surface (PES) for a collection of N atoms can be considered as a multidimensional hypersurface that is dependent on the resulting 3N-6 internal degrees of freedom associated with that system. An energy change occurs when any of the degrees of freedom are altered and for most systems, it is not usually viable, or necessary to investigate the entire hyper-surface. Subsequently, an energy landscape can be generated from the systematic alteration of, say, two independent geometrical parameters. As the value of each parameter is changed, the energy is minimised in all the other degrees of freedom. This process allows a thorough investigation into a region of the hypersurface associated with the chemical change under investigation.

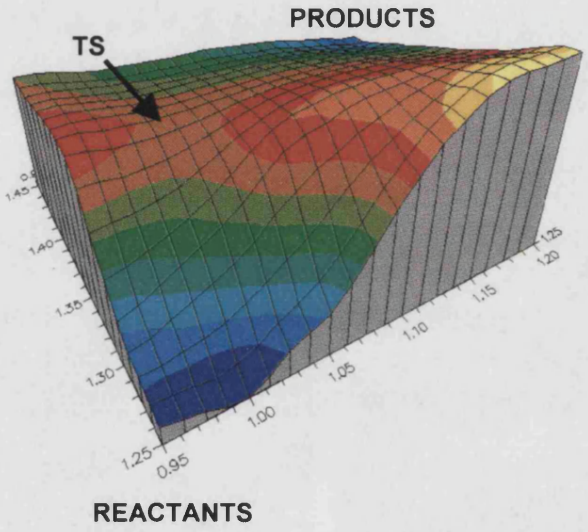


Fig 2.1. Example of an energy landscape produced by the systematic alteration of two degrees of freedom. Blue corresponds to the lowest energy points on the surface and red shows the highest energy points.

On the surface, there exist a number of stationary points. Minima usually correspond to equilibrium geometries, such as reactants and products on a reaction profile. The route taken across the surface for any reaction is the lowest energy pathway possible, with reactants and products located at minima on either side of the surface. For a chemical step such as that depicted in figure 2.1, there is a maximum on the lowest energy pathway corresponding to a *first-order saddle point*. The structure at such a

point is known as the *transition structure* (TS) and it determines the energy barrier that must be overcome when proceeding from reactants to products. There also exist higher-order saddle points, but this thesis is only concerned with the determination of the chemically significant TS.

For any stationary point on the PES, the energy change with respect to all the variables (i.e. the force) is zero. For *i* variables therefore:

$$\left(\frac{\partial E}{\partial R_1}, \frac{\partial E}{\partial R_2}, \frac{\partial E}{\partial R_3}, \frac{\partial E}{\partial R_4}, \frac{\partial E}{\partial R_5}, \ldots\ldots\ldots\ldots \frac{\partial E}{\partial R_i}\right) = 0 \qquad (2.1)$$

This verifies the presence of a stationary point on the PES; however, the nature of such a point (i.e. whether it is a minimum or maximum) is determined through the knowledge of the surface curvature. This is defined by the second derivatives of the energy and leads to an *i* x *i* matrix known as the *hessian*:

$$\begin{pmatrix} \frac{\partial^2 E}{\partial R_1^2} & \frac{\partial^2 E}{\partial R_1 \partial R_2} & \frac{\partial^2 E}{\partial R_1 \partial R_3} & \frac{\partial^2 E}{\partial R_1 \partial R_4} & \cdots & \frac{\partial^2 E}{\partial R_1 \partial R_i} \\ \frac{\partial^2 E}{\partial R_2 \partial R_1} & \frac{\partial^2 E}{\partial R_2^2} & \frac{\partial^2 E}{\partial R_2 \partial R_3} & \frac{\partial^2 E}{\partial R_2 \partial R_4} & \cdots & \frac{\partial^2 E}{\partial R_2 \partial R_i} \\ \frac{\partial^2 E}{\partial R_3 \partial R_1} & \frac{\partial^2 E}{\partial R_3 \partial R_2} & \frac{\partial^2 E}{\partial R_3^2} & \frac{\partial^2 E}{\partial R_3 \partial R_4} & \cdots & \frac{\partial^2 E}{\partial R_3 \partial R_i} \\ \frac{\partial^2 E}{\partial R_4 \partial R_1} & \frac{\partial^2 E}{\partial R_4 \partial R_2} & \frac{\partial^2 E}{\partial R_4 \partial R_3} & \frac{\partial^2 E}{\partial R_4^2} & \cdots & \frac{\partial^2 E}{\partial R_4 \partial R_i} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 E}{\partial R_i \partial R_1} & \frac{\partial^2 E}{\partial R_i \partial R_2} & \frac{\partial^2 E}{\partial R_i \partial R_3} & \frac{\partial^2 E}{\partial R_i \partial R_4} & \cdots & \frac{\partial^2 E}{\partial R_i^2} \end{pmatrix} \qquad (2.2)$$

Under the harmonic approximation (i.e. a symmetrical potential energy well), the force, F_i, on a given geometrical variable, R_i, is given by:

$$F_i = \kappa_i \Delta R_i \qquad (2.3)$$

where κ_i is the force constant. If R_i represents a bond, then the force constant is a measure of the bond strength. A high value for the force constant represents a stiff bond (with a narrow potential) while a smaller force constant results in a weaker, broader potential (figure 2.2). The second derivative of the energy with respect to the geometrical parameters leads to the force constants. For a diatomic molecule, the vibrational frequency, ν, associated with the bond stretch is given by:

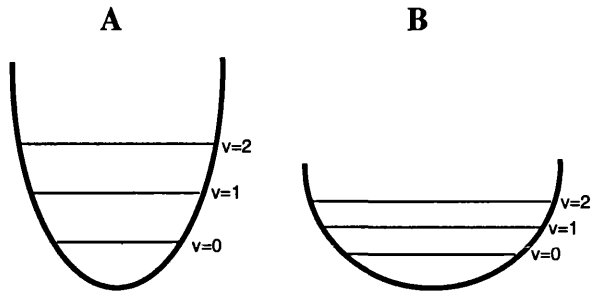


Fig 2.2. Schematic diagram showing a). a tight potential (large force constant) and b). a loose potential (small force constant). Both situations assume a harmonic potential; this is appropriate for the base of the potential energy well.

$$\nu = \frac{1}{2\pi}\sqrt{\frac{\kappa}{\mu}} \tag{2.4}$$

where μ is the reduced mass of the molecule. For an N atom system, the hessian matrix in Cartesian co-ordinates has $3N$ by $3N$ elements. The vibrational frequencies are obtained from the eigenvalues of the mass-weighted hessian and the vibrational modes are the corresponding eigenvectors. For a non-linear molecule (located at a stationary point), there are 6 zero eigenvalues corresponding to the overall translational and rotational motion of the molecule in three dimensional space.

For minima stationary points, the energy is a minimum in all directions; the resulting eigenvalues of the hessian are all positive and the vibrational frequencies have real values. The first-order saddle point, however, is a maximum in one direction and a minimum in all the others. At this point, therefore, one (and only one) eigenvalue in the hessian is negative and the associated vibrational frequency is imaginary (since the square root of the negative eigenvalue is used to evaluate the frequency).

The first stage of modelling chemical reactivity must involve the generation of the PES.

2.2. Generation of the Potential Energy Surface

2.2.1. Quantum Mechanics[3,5,6]

An electron has wave-like properties whose nature is determined by its kinetic energy and interaction with other particles in the system; such properties are described by the Schrödinger equation:

$$H\Psi=E\Psi \tag{2.5}$$

H is the Hamiltonian that operates upon the wavefunction and consists of terms that represent inter-particle interactions as well as the kinetic energy of the electron. For a molecular system that contains N electrons as well as M nuclei with charge Z, the Hamiltonian (in atomic units) is given by:

$$H = \sum_{i=1}^{N} -\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{\alpha=1}^{M}\frac{Z}{r_{i\alpha}} + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\frac{1}{r_{ij}} + \sum_{\alpha=1}^{M}\sum_{\beta=\alpha+1}^{M}\frac{Z_\alpha Z_\beta}{r_{\alpha\beta}} \tag{2.6}$$

where,

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{2.7}$$

The first term in eqn. 2.6 is the kinetic energy of the electrons while the second, third and fourth terms represent electron-nuclear, electron-electron and nuclear-nuclear interactions respectively. The kinetic energy of the nuclei are not considered here: it is assumed that electrons interact with a particular configuration of stationary nuclei (whose mass is *ca.* 2000 times greater than an electron). This is known as the *Born-Oppenheimer* approximation and it renders the fourth term in eqn. 2.6 a constant for a particular nuclear configuration.

The probability of finding an electron in space is taken to be the square of the wave function; this gives the electron density at a certain point. If integrated over all space, the probability of finding the electron of interest must be equal to one:

$$\int \psi^* \psi \, d\tau = 1 \tag{2.8}$$

When this condition is true, the wavefunction is *normalised.* Solutions to the Schrödinger equation are also required to adhere to this restriction. Thus considering eqn. 2.5 in terms of electron probabilities gives:

$$\int \psi^* H \psi \, d\tau = E \int \psi^* \psi \, d\tau \tag{2.9}$$

Thus for a normalised wavefunction:

$$E = \int \psi^* H \psi \, d\tau \tag{2.10}$$

The *one-electron* Hamiltonian, h, for electron, i, in the molecular system is given by its kinetic energy and interactions with all nuclei:

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_{\alpha=1}^{M} \frac{Z_\alpha}{r_{i\alpha}} \tag{2.11}$$

Each electron is contained within one spin-molecular orbital, χ_i (where there are two spin molecular orbitals to a spatial molecular orbital). Using eqn. 2.10, the energy of each individual electron is given by:

$$E = H_{ii}^{core} = \int \chi_i(h_i)\chi_i \, d\tau \tag{2.12}$$

where, h_i has been defined in eqn. 2.11. H_{ii}^{core} represents the energy of the electron in a bare field of nuclei. In addition to these one-electron integrals, *two-electron integrals* account for the repulsion energy between electrons. Thus, the repulsion between electron 1 in spin orbital, χ_i, and electron 2 in spin orbital, χ_j, is:

$$E = J_{ij} = \int\int \chi_i(1)\chi_j(2)\left(\frac{1}{r_{12}}\right)\chi_i(1)\chi_j(2) d\tau_1 d\tau_2 \tag{2.13}$$

Between two electrons with like-spin, however, there exists a small favourable energy term: electrons with the same spin are not permitted to occupy the same position in space, as stated by the *Pauli exclusion principle.* The subsequent reduced

repulsion between electrons with the same spin-orientation manifests itself as favourable energy contribution and this is known as an *exchange interaction*:

$$E = K_{ij} = \int\int \chi_i(1)\chi_j(2)\left(\frac{1}{r_{12}}\right)\chi_j(1)\,\chi_i(2)d\tau_1 d\tau_2 \qquad (2.14)$$

For a closed shell system, two electrons in one spatial orbital interact in four different ways with two electrons in another spatial orbital (figure 2.3).

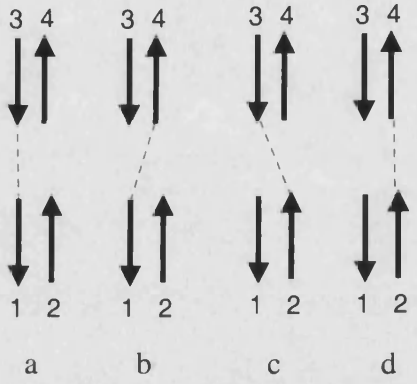


Fig 2.3. Various ways by which two electrons in one spatial orbital can interact with the electrons in another.

The total repulsion energy between the two pairs of electrons in figure 2.3 is thus $4J_{ij}$; this is an over-estimate of the coulombic interactions because two of the situations involve unpaired spins (figure 2.3a, figure 2.3d). This necessitates inclusion of the exchange energy ($2K_{ij}$). Finally, electron-electron interactions, J_{ii}, within the same spatial orbital must also be taken into account. Thus for a closed shell system with N/2 filled spatial molecular orbitals, the total electronic energy of the system is:

$$E = 2\sum_{i=1}^{N/2} H_{ii}^{core} + \sum_{i=1}^{N/2}\sum_{j=i+1}^{N/2}(4J_{ij} - 2K_{ij}) + \sum_{i=1}^{N/2} J_{ii} \qquad (2.15)$$

Because $J_{ii}=K_{ii}$, this equation can be simplified by "double counting" all the electron-electron interactions found within the system:

$$E = 2\sum_{i=1}^{N/2} H_{ii}^{core} + \sum_{i=1}^{N/2}\sum_{j=1}^{N/2}(2J_{ij} - K_{ij}) \qquad (2.16)$$

This is the *Hartree-Fock* (HF) electronic energy for a closed shell system with N electrons occupying $N/2$ molecular orbitals. It should be noted that the overall HF energy would also account for the (constant) nuclear-nuclear repulsions.

The Hartree-Fock approach in determining the energy of the system involves altering the wavefunction for each individual electron within a fixed electronic configuration. This wavefunction is then fixed and the next orbital in the system is freed and altered to lower the molecular energy. This process is done for all N electrons and subsequently repeated until the energy of the system can't be lowered any further. At this point, the molecular orbitals are said to be *self-consistent.*

The spatial molecular orbitals, ψ_i, are formulated by *Linear Combination of Atomic Orbitals* (LCAO). Thus, T atomic orbitals, ϕ, contribute to forming T molecular orbitals:

$$\psi_i = \sum_{j=1}^{T} c_j \phi_j \tag{2.17}$$

The constant, c, determines the magnitude of contribution that atomic orbital, ϕ_j makes to the eventual molecular orbital ψ_i. The atomic orbitals are also known as *basis functions*, and a complete description of the molecular orbital is obtained from an infinite number of basis functions (where T=∞). Therefore, the more basis functions used to generate the molecular orbitals, the more complete the description of the system.

The formation of self-consistent molecular orbitals is therefore achieved by altering the coefficients of each atomic orbital iteratively. The best possible wavefunction for the molecular system is defined as the eigenfunction that results in the lowest possible energy solution to the Schrödinger equation (eqn 2.5). The number and type of basis functions to be used must be supplied by the user in the form of basis sets.

2.2.1.1. Basis Sets

The basis sets used within this thesis consist of a set of gaussian functions (gf) of the form, $\exp(-\alpha r^2)$. The α term represents how broad the gf is, with smaller values giving a more diffuse function. Each basis function comprises of a linear combination of gaussian functions: the use of individual gaussians provides a poor description of the electron density at the nucleus and at longer electron-nuclear distances. Thus:

$$\phi_i = \sum_{j=1}^{G} d_j \gamma_j \qquad (2.18)$$

where, γ_j are the gaussian functions used in generating atomic orbital, ϕ_i, and d_j is the corresponding coefficient for each gf. To generate self-consistent molecular orbitals, the values of d_j and α are altered throughout the calculation for every gf in the basis set. The labelling of such basis sets is as follows:

W-XYZ (Dif1, Dif2) G(Pol1, Pol2)

W represents the number of gaussian functions describing the "inner" electrons (i.e. the 1s electrons for second row elements); for all basis sets used herein, W=6. The X, Y and Z basis functions describe the valence shell electrons, in which the X function is contracted while Y and Z are more diffuse. This allows increased orbital flexibility when minimising the energy of the system compared to only having one basis function alone. The *split-valence double-zeta* basis set describes the valence shell orbitals using two basis functions (namely, X and Y). The basis sets used for the work in this thesis usually have X=3 and Y=1. Occasionally, the more flexible *split-valence triple-zeta* basis set is used in which Z is generated from one gf (i.e. Z=1).

Because electron density away from the nucleus is underestimated when using gaussian type orbitals, anions need a further set of diffuse s- and p-functions for an appropriate description of their valence electrons. Dif1 adds diffuse functions to all

heavy atoms in the system and is designated with a "+". Further addition of diffuse functions to hydrogen atoms (Dif2) is designated "++".

The charge distribution in an atomic orbital is regarded to be isotropic, but is usually not symmetrical when forming a molecular orbital. To account for this, the charge distribution in atoms needs to be polarised. This is achieved by augmenting valence shell orbitals with functions of higher azimuthal quantum number (i.e. s-orbitals are combined with p-functions and p-orbitals are combined with d-functions), for example:

Fig 2.4. Polarisation of a p-orbital by a d-function.

Polarisation functions in the basis set are designated after the "G" with "p", "d" or "f", depending on the variety used. Pol1 and Pol2 represent sets of polarisation functions on heavy and light atoms respectively. Thus, for example (d,p) adds d-functions to heavy atoms and p-functions to hydrogens.

2.2.1.2. Electron Correlation

The HF energy underestimates the correlated motion of electrons and thus overestimates the 2-electron repulsions in the system, giving rise to a species with higher energy than the "true value". This shortcoming can be accounted for by using *Møller-Plesset perturbation theory* (MPPT). This involves taking the Hamiltonian of the solvable N one-electron system, H_o (eqn. 2.11), and perturbing it by λ to give the exact Hamiltonian of the system, such that:

$$H_{exact} = H_o + \lambda H \qquad (2.19)$$

The exact Hamiltonian and energy can be subsequently expanded into a power series:

$$H_{exact} = H_o + \sum_n \lambda^n H_n \tag{2.20}$$

$$E_{exact} = E_o + \sum_n \lambda^n E_n \tag{2.21}$$

where n is the number of members in the power series and λ is raised to the power n. H_n is the nth order perturbation of H_o and E_n is the corresponding nth order energy correction for the exact energy. The first order correction represents the energy associated with the two-electron repulsion integrals. Thus, the HF energy is given by $E+E_1$. The second energy correction, E_2, is the first improvement of the Hartree-Fock energy and is obtained from a *second-order Møller-Plesset perturbation* calculation (MP2).[7] In the Gaussian suite of programs,[8] the MP2 calculation can be done with inclusion of the 1s electrons ("full") or exclusion of them ("frozen core", fc). MP2 is size consistent, but doesn't account for electron excitations systematically.

The other method used to account for electron correlation effects at *ab initio* levels of theory in this thesis is the more computationally demanding *configuration interaction* (CI). This involves incorporating different electron configurations into the closed-shell ground-state wavefunction through systematic electron excitations into the ground-state virtual orbitals. This results in a "better" wavefunction that leads to a system with lower energy, hence:

$$\psi_z = c_o\psi_o + \sum_s c_s\psi_s + \sum_d c_d\psi_d + \sum_t c_t\psi_t + \sum_q c_q\psi_q + ... \tag{2.22}$$

The better wavefunction, ψ_z, can be given as a linear combination of wavefunctions associated with singlet (ψ_s), doublet (ψ_d), triplet (ψ_t) and higher orders of electron excitations. The hydrogen molecule, for example, has only two electrons; including single and double excitations into the molecular description (using a finite basis set) thus gives *full configuration interaction* for H_2. CI calculations that involve both single and double excitations are referred to as CISD.

One complication of CI is that the method is not size-consistent. Thus, for example, two infinitely separated helium atoms would have full CI using CISD. However, an

He_2 complex described at this level of theory would exclude triplet and quadruplet excitations necessary to fully account for CI in the four-electron system. However, Pople *et al.* have added a quadratic term to the CI method that makes such calculations size-consistent; this version of CI is known as *Quadratic Configuration Interaction* (QCI).[9] Pople *et al.* went on to formulate the QCISD method and then approximated the term associated with triplet excitations by using a perturbation method. This is known as QCISD(T).[9] The triplet approximation allows for greater computational efficiency than systematically accounting for all triplet excitations.

2.2.1.3. The AM1 Semi-empirical Method

Ab initio techniques involve evaluation of many two-electron repulsion integrals and such methods are thus computationally expensive. To consider a less computationally demanding technique must therefore require the calculation of a reduced number of integrals. AM1 (Austin Model 1)[10] is based on the Neglect of Diatomic Differential Overlap (NDDO). If two electrons are described by two basis functions, then the two-electron integral is:

$$\int\int \phi_a(1)\phi_b(1)(\frac{1}{r_{12}})\phi_c(2)\,\phi_d(2)d\tau_1 d\tau_2 \qquad (2.23)$$

where basis functions *a* and *b* are used to describe electron 1 and basis functions *c* and *d* describe electron 2. If *a* and *b* are on different centres (Figure 2.5c), then this integral is set to zero under the AM1 technique. This is also the case if *c* and *d* are located on different centres and is in contrast to *ab initio* levels of theory that do calculate these three- and four-centre repulsion integrals. The AM1 method thus evaluates the two-centre two-electron integrals (figure 2.5a), the one-centre two-electron integrals (figure 2.5b), the core-electron attraction integrals and the core-core repulsions. To allow rapid evaluation of these integrals, AM1 is dependent on a set of experimentally determined parameters (e.g. ionisation energies) for each atom type. In addition to calculating a smaller number of integrals, the AM1 method incorporates the 1s electrons and nucleus into a single core. The rationale for this approximation is that important chemical changes involve only the valence electrons.

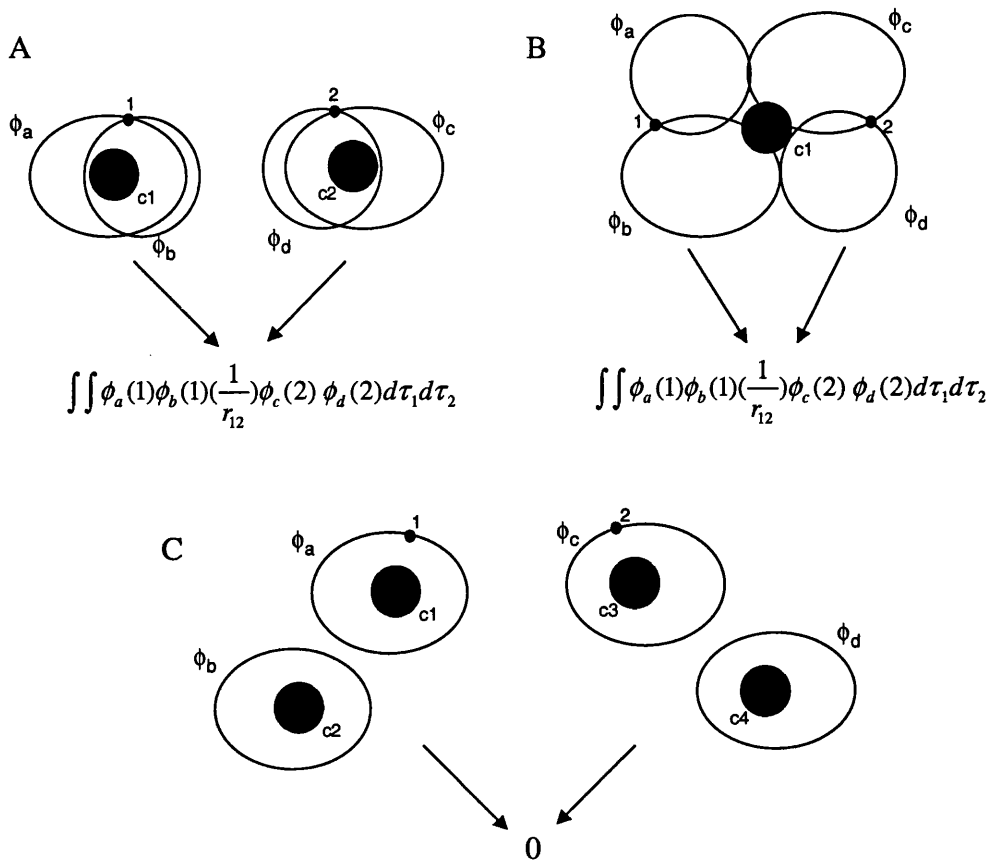


Fig 2.5. Electron repulsion integrals are calculated at AM1 for a) two-centre, two-electron and b) one-centre, two-electron interactions. The AM1 method however omits two-electron repulsions over three and four centres, the latter shown in C.

2.2.2. Molecular Mechanics[3]

Unfortunately, even semi-empirical QM methods are currently too computationally demanding to model systems with more than 200 atoms or so. Enzymes and explicit solvation models that consist of many thousands of atoms are therefore unable to be described at this level of theory. Thus a different method is required to model these systems.

The approach taken in describing large molecular systems is to use a simpler approximation: to model the system at the atomic level rather than at the sub-atomic stage. To do this, molecular mechanics programs such as CHARMM[11] describe the system in terms of balls (atoms) connected by harmonic springs (bonds). The MM *force field* thus consists of force constants and structural parameters for several types

of atom; it relates the energy of the system to its structure using geometrical parameters such as bond-stretches, angle-bends and angle-torsions:

$$E_{struct} = \sum_{bonds} \frac{k_b}{2}(l - l_o)^2 + \sum_{angles} \frac{k_a}{2}(\theta - \theta_o)^2 + \sum_{torsions} \frac{k_t}{2}[1 - cos(n\omega)] + \sum_{impropers} \frac{k_i}{2}(\omega - \omega_o)^2 \qquad (2.24)$$

where l, θ, and ω are the values for bond lengths, angles and torsions respectively. The last term accounts for out-of-plane torsions and is included to maintain planarity about planar centres (e.g. carbonyl systems). The reference values l_o, θ_o, and ω_o are equilibrium geometrical parameters in the absence of any other term in the force field. These values are supplied as parameters and in the case of the CHARMM force field have been determined from spectroscopic and X-ray crystal data.[12,13] In addition, the related force constants for each "spring" are also supplied and are usually determined using infrared and Raman spectroscopic data and/or high-level *ab initio* techniques.[12,13] Further to these parameters, a value for the torsional folding term, *n*, must also be supplied for each atom type depending on the symmetry of the centre. Thus, a methyl group in ethane, for example, should have three minimum conformations (two gauche and one staggered), i.e. n=3.

In addition to the structural terms, there are also *non-bonded* terms that include electrostatic interactions and van der Waals forces:

$$E_{non-bond} = \sum_{i=1}\sum_{j=1} \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} + \sum_{i=1}\sum_{j=1}\left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}}\right) \qquad (2.25)$$

The first term in eqn. 2.25 is the coulombic interaction between two atom-centred partial charges;[12,13] these are summed over all particles in the system that are not directly bonded to one another. The second term is the van der Waals interaction between species, *i*, and species, *j*, and is often referred to as the *Lennard-Jones 12-6* potential (L-J 12-6). The total energy of the system is usually the summation of the terms in eqn. 2.24 and eqn. 2.25. The majority of computation time associated with MM calculations is due to calculation of the non-bonded interactions.

In CHARMM, it is also possible to constrain atoms using rigid constraints or with a harmonic constraint of the form:

$$E_{harm} = \sum_{h} k_h (r_p - r_{ref})^2 \qquad (2.26)$$

where the starting co-ordinate for atom *h* is designated r_{ref} and its position in the system under study is r_p. The force constant, k_h, (which can be mass weighted) is supplied by the user and the resulting energy is added to the total energy of the system.

The simulations in this thesis use an *all atom* force field that explicitly includes all atoms in the molecular system.[13] However, when computational resource is not available, it is also possible to use a *united atom* force field. This incorporates some or all of the hydrogens into the heavy atoms to which they are bonded. Consequently, the resulting "super-atoms" require a different set of parameters compared to the *all atom* force field. Such descriptions, however, can not accurately model hydrogen bonding and steric effects associated with individual hydrogen atoms.

2.2.3. Quantum Mechanics / Molecular Mechanics (QM/MM) Hybrid Modelling

Molecular mechanics provides a way of modelling many thousands of atoms by using an atomic description of the system. The drawback of using such methods is that electronic description in the system is ignored. Since chemical reactivity, by inference, requires the description of the re-organisation of electrons in the system, MM ordinarily is unable to model chemical processes. However, this is the strength of the computationally intensive QM methods. Modelling chemical reactions within biological systems such as enzymes could thus be achieved by combining the two techniques. Such a method currently exists in CHARMM; the version used in this thesis uses the semi-empirical AM1 level of theory (originally from MOPAC) in conjunction with the CHARMM all atom force field.[14]

The QM/MM system in CHARMM is divided into three sections (figure 2.6): A QM region describes a sub-set of atoms located at the reacting centre while the MM part of the system includes atoms around the QM region.

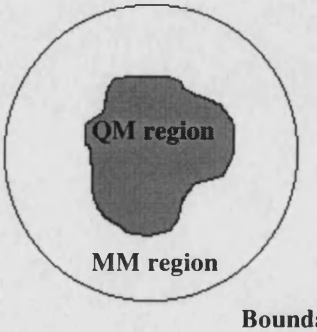


Fig 2.6. The partitioning in a QM/MM system.

Finally, atoms beyond explicitly described system can be modelled using boundary techniques, such as the stochastic boundary method of Brooks and Karplus.[15]

The effective Hamiltonian for the resulting QM/MM system may be given by:

$$H_{eff} = H_{QM} + H_{QM/MM} + H_{MM} + H_{bound} \tag{2.27}$$

H_{QM} is the QM Hamiltonian incorporating all sub-atomic particle interactions in the QM part and is given in equation 2.6 for an N electron and M nuclei system. The QM/MM Hamiltonian represents a cross term in which the QM part is affected by the MM system and is given by:

$$H_{QM/MM} = -\sum_{i=1}^{N}\sum_{m=1}^{X}\frac{Q_m}{r_{im}} + \sum_{\alpha=1}^{M}\sum_{m=1}^{X}\frac{Z_\alpha Q_m}{r_{\alpha m}} + \sum_{\alpha=1}^{M}\sum_{m=1}^{X}\left(\frac{A_{\alpha m}}{r_{\alpha m}^{12}} - \frac{B_{\alpha m}}{r_{\alpha m}^{6}}\right) \tag{2.28}$$

The first term accounts for the interactions between N electrons and X MM atoms with point charge, Q_m in a QM/MM system. The term relies upon the co-ordinates of the electrons in the QM region and must therefore be included into the evaluation of the electronic wavefunction. The second term is the interaction between the QM nuclei (with charge Z) and the MM atoms (with charge Q_m) separated by distance r. Once evaluated, this term is a constant as it is not dependent on the positions of the QM electrons; in CHARMM, it is evaluated by placing notional s-orbitals at each MM centre. The L-J 12-6 potential accounts for the van der Waals interaction between atoms in the MM and QM partitions; it is described by the relevant non-bonded parameters in the MM force field. The inclusion of the L-J potential allows the QM system to differentiate between like-charged, different size species (*cf.* bromide and chloride anions); omission of this term would make any size dependent non-bonded interactions invisible to the QM region.

The last two terms in eqn. 2.27 are H_{MM} and H_{bound}. The H_{MM} represents the energy of the MM part of the system, while H_{bound} describes the effect of the boundary on the explicitly described atoms. The total energy of the QM/MM system is thus given by the combination of the separate QM, QM/MM, MM and boundary energies:

$$E_{tot} = E_{QM} + E_{QM/MM} + E_{MM} + E_{bound} \quad (2.29)$$

In considering an explicitly solvated solute, there exists a natural QM and MM partition: the solute can be treated quantum mechanically while the surrounding solvent molecules are described using a molecular mechanics potential. However, large molecular systems such as enzymes are frequently involved with the reaction directly. It is currently not feasible to treat the entire enzyme quantum mechanically; therefore, a single molecular entity must be described using both QM and MM techniques.

This is achieved in CHARMM by applying the *link atom approximation.* This first involves a decision as to which enzyme residues are directly involved in the reaction. These residues need to be treated quantum mechanically to allow an electronic description at the reaction centre. However, because the remainder of the enzyme is treated with the simpler MM description, an extra QM atom must be applied to satisfy the valency requirements of the QM system. The quantum link atom used in CHARMM is a hydrogen atom. In figure 2.7, for example, the side-chain of an active site aspartate residue is treated quantum mechanically. A QM link atom is placed along the C_α—C_β bond about 1.0Å away from the QM described C_β atom. Thus, the valency of the C_β atom is satisfied and the Asp side chain is modelled by an acetate anion in the QM system. The QM link atom has no non-bonded or internal co-ordinate parameters associated with it and so is not included in the MM part of the calculation. However, the QM link atom is fully visible to other QM atoms and its description is thus included in H_{QM}.

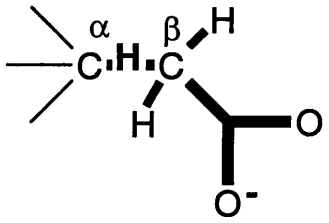


Fig 2.7. QM link atom (**H**) addition to an enzyme Asp residue between C_α and C_β. This yields an "acetate" QM system.

In the QM/MM system, any MM internal co-ordinate terms (eqn. 2.24) associated with QM atoms alone are deleted. At the QM and MM junction however, such terms for the QM atoms remain and thus allow a full classical description of the environment around the junction MM atom(s).

2.2.4. Continuum Solvation Techniques

While it is usually desirable to incorporate specific solvent-solute interactions within a simulation, especially if a protic-polar solvent is being modelled, computational requirements for such descriptions can be high. An alternative approach is to surround the solute with a solvent description in the form of a continuum with dielectric constant, ε. Two continuum methods are used within this thesis to model bulk solvent water (ε=78.4).

The continuum solvation method used with *ab initio* levels of theory was chosen to be the *isodensity polarised continuum model* (IPCM) of Frisch and co-workers.[16] With this method, the continuum is modelled using a surface of even electron density that reasonably defines the molecular shape of the solute. The dipole and multipoles of the solute interact with the surrounding dielectric and results in the formation of a dipole in the continuum; this is reflected back on to the solute giving it greater stability. If the electron density of the solute subsequently alters, then there is a corresponding change to the isosurface. The Hamiltonian for a solute immersed in such a continuum usually includes three terms. These are the gas phase Hamiltonian of the solute; the interactions between the isosurface and the solute electrons; and the interactions between the isosurface and solute nuclei. For the implementation used within this thesis, no geometry optimisation takes place within the continuum. Thus, the calculations give the energy for a solvated *in vacuo* geometry.

The continuum model used with the AM1 method is the *conductor-like screening model* (COSMO).[17] Its implementation is somewhat different to the IPCM method. Each atom in the solute is considered as a point charge within a solvent cavity. The continuum cavity is evaluated by taking the radius of each solute atom and then subsequently adding the effective radius of the solvent molecules to it (figure 2.8). The solvent radius is supplied as a parameter and is the effective radius of the solvent molecule. The solute atoms form a dipole within the continuum that subsequently lowers the energy of the solute. The calculation of such effects is done by splitting the continuum spheres up into smaller segments. The dipole induced at

each segment of the continuum interacts with dipoles on the other segments; by minimising these dipole-dipole interactions the lowest energy for the solvated solute can be evaluated.

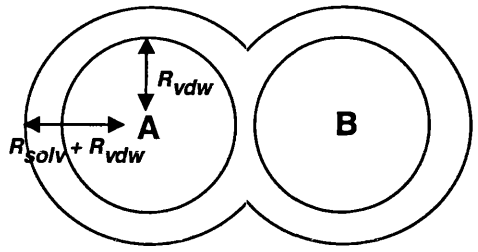


Fig 2.8. Two atoms, *A* and *B* within a molecule surrounded by a continuum dielectric as described by COSMO. The solvent cavity associated with each solute atom is given by $R_{solv}+R_{vdw}$ from the atom centre. R_{vdw} is the van der Waals radius of the atom. R_{solv} is the effective radius of solvent molecules.

Unlike the IPCM method, COSMO calculations allow full optimisation of solute immersed within the continuum dielectric.

2.3. Exploration of the Potential Energy Surface

2.3.1. Simulated Annealing[3]

Large molecular systems such as enzymes and explicitly described solvent environments result in a very large conformational space. Ideally, this needs to be sampled in order to find the lowest energy arrangements of the system, but this is unlikely to be achieved systematically on a reasonable time scale.

An approach used for one of the studies in this thesis is *simulated annealing*. This involves heating the system to a specified temperature and then allowing it to reach equilibrium before undergoing controlled cooling. As the system is cooled, lower energy-states become more probable. Thus, the lowest energy-state should be obtained at absolute zero. However, controlling the temperature during simulated annealing is difficult and the system often drops into a local minimum with an energy barrier that is greater than the thermal energy contained within the system (figure 2.9).

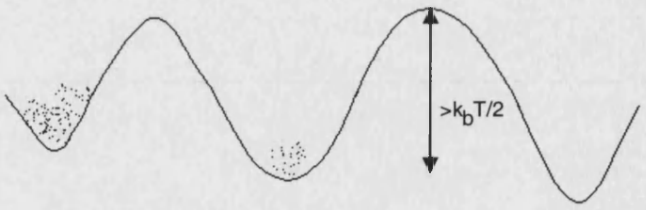


Fig 2.9. Lower energy-states begin to be occupied as the temperature of the system is dropped. However, if the barrier is larger than ½ k_bT (where k_b is the Boltzmann constant), then occupancy of a lower energy-state is not classically probable.

Simulated annealing experiments in this thesis have been done using a molecular dynamics method. Such simulations require numerical differentiation because the forces acting on a single entity are dependent on all the other species within the system. The method used in this thesis is the *Leapfrog* algorithm. This divides the dynamics into a series of time frames, δt and first calculates the velocity at $(t + \frac{1}{2}\delta t)$:

$$\mathbf{v}(t + \tfrac{1}{2}\delta t) = \mathbf{v}(t - \tfrac{1}{2}\delta t) + \delta t\mathbf{a}(t) \qquad (2.30)$$

where $\mathbf{a}(t)$ is the acceleration at time *t*. Once $\mathbf{v}(t + \frac{1}{2}\delta t)$ has been evaluated, the new atomic positions can be calculated:

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t + \tfrac{1}{2}\delta t) \qquad (2.31)$$

The initial velocities of the atoms within the system are randomly generated with a gaussian distribution in all the simulated annealing experiments presented in this thesis; the time step chosen was one femtosecond.

2.3.2. Intrinsic Reaction Co-ordinate

Another way of exploring the PES is to trace the lowest energy pathway for a chemical process. Once the TS has been determined, a path of steepest descents can be followed (in the direction of the mode associated with the negative eigenvalue) to the reactant and product valleys. Fukui has devised such an approach that considers steps along the reaction path in terms of nuclear motions with infinitesimal

velocity.[18] Thus at each stage, the kinetic energy along the reaction pathway is near zero. This *intrinsic reaction co-ordinate* (IRC) approach allows verification of the local reactants and products that the TS connects and this is essential when calculating any isotope effects for a chemical process.

2.4. Analysis of the Potential Energy Surface

2.4.1. Isotope Effects[4]

Theoretically modelling transition structures can lead to the calculation of experimentally observable properties, namely, the kinetic and equilibrium isotope effects (KIE and EIE respectively). Such calculations can therefore allow full interplay between theory and experiment, in which theory is able to give insight into the experimentally non-determinable TS structure. Isotope effects are obtained by substituting a specific atom for a heavier isotope.

Because a neutron has a unit charge of zero, its subsequent addition to a nucleus doesn't alter the electronic description of the system under the Born-Oppenheimer approximation. Such a change thus results in no alteration of the PES or its curvature. The force constants contained in the hessian therefore have the same value for an isotopically substituted system as compared with non-isotopically substituted case. However, the vibrational frequencies are altered because the mass (and hence the reduced mass) of the molecule is increased (*cf.* eqn 2.4). The zero-point vibrational energy for a particular vibration (e.g. bond stretch) is reduced with a heavier isotope; for an isotopically substituted system therefore, a greater amount of energy is required to dissociate the molecule.

The individual populations of i energy levels for a particular system vibration can be given by the Boltzmann distribution law:

$$\frac{n_i}{n_o} = exp-\{(\varepsilon_i - \varepsilon_o)/k_B T\} \qquad (2.32)$$

n_i is the number of molecules found in energy level i with n_o being the population of the ground vibrational state. The corresponding energies for each energy level are given by ε_i and ε_o (figure 2.10). It follows that the summation of all the individual energy level populations, Σn_i, can be related to the total number of particles in the system by:

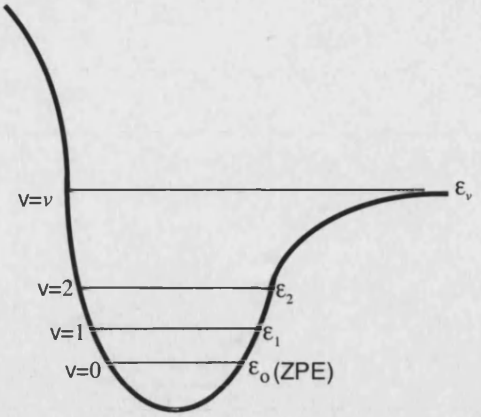


Fig 2.10. Potential curve for one degree of freedom within a molecular system indicating the quantised energy levels.

$$N = n_o \sum_{i=1}^{v} exp-\{(\varepsilon_i - \varepsilon_o)/k_B T\} \qquad (2.33)$$

The summation is known as the partition function, Q, and represents the distribution of N species within v energy levels for a particular degree of freedom (i.e. $N=n_oQ$).

For two species in equilibrium in which the ZPE levels are separated by $\Delta\varepsilon_o$, the equilibrium constant is:

$$K = \frac{N_a}{N_b} = \frac{Q_a}{Q_b}.exp-\{\Delta\varepsilon_o / k_B T\} \qquad (2.34)$$

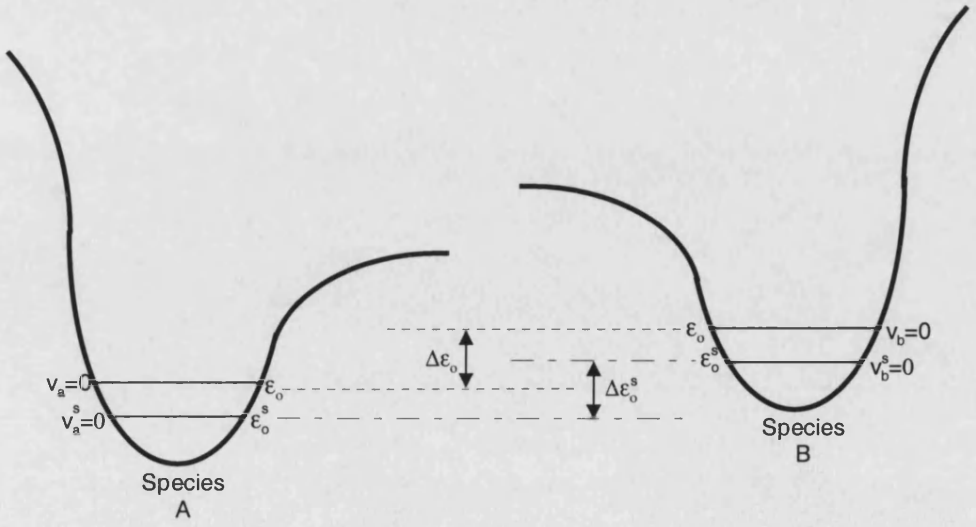


Fig 2.11. Schematic diagram showing the differences in the ZPE along the "reaction co-ordinate" degree of freedom for species A and B in equilibrium.

The resulting energy difference between the ZPE levels for isotopically substituted *A* and *B* species is likely to be different as compared to the unsubstituted molecules. This leads to a different the equilibrium constant, K^s, for the substituted system:

$$K^s = \frac{N_a^s}{N_b^s} = \frac{Q_a^s}{Q_b^s}.exp-\{\Delta\varepsilon_o^s / k_B T\} \qquad (2.35)$$

The resulting *equilibrium isotope effect* is presented as a K/K^s fraction and is given by:

$$\frac{K}{K^s} = \frac{Q_a}{Q_b}.\frac{Q_b^s}{Q_a^s}.exp-\{(\Delta\varepsilon_o - \Delta\varepsilon_o^s)/k_B T\} \qquad (2.36)$$

where $\Delta\varepsilon_o$ and $\Delta\varepsilon_o^s$ are defined in figure 2.11. A similar equation can be derived for the *kinetic isotope effect* (KIE) except state *B* is now located at a first-order saddle point (and thus labelled *TS*). The resulting rate constant ratio is given by:

$$\frac{k}{k^s} = \frac{Q_a}{Q_{TS}}.\frac{Q_{TS}^s}{Q_a^s}.exp-\{(\Delta\varepsilon_o - \Delta\varepsilon_o^s)/k_B T\} \qquad (2.37)$$

Where $\Delta\varepsilon_o$ and $\Delta\varepsilon_o^s$ are now zero-point vibrational activation energies for the unsubstituted and substituted processes respectively and are shown in figure 2.12. In addition to the partition functions associated with vibrations, there will also be partition functions for rotation and translation motions as well. These are sometimes referred to as mass and moments of inertia (MMI). Usually, the product of the MMI and vibrational partition functions in eqn. 2.37 is close to unity and the isotope effect can be given by just the exponential term as a fair approximation. In this thesis, the program CAMISO[21] used to evaluate isotope effects doesn't make this approximation and evaluates both MMI and the vibrational partition functions of the system.

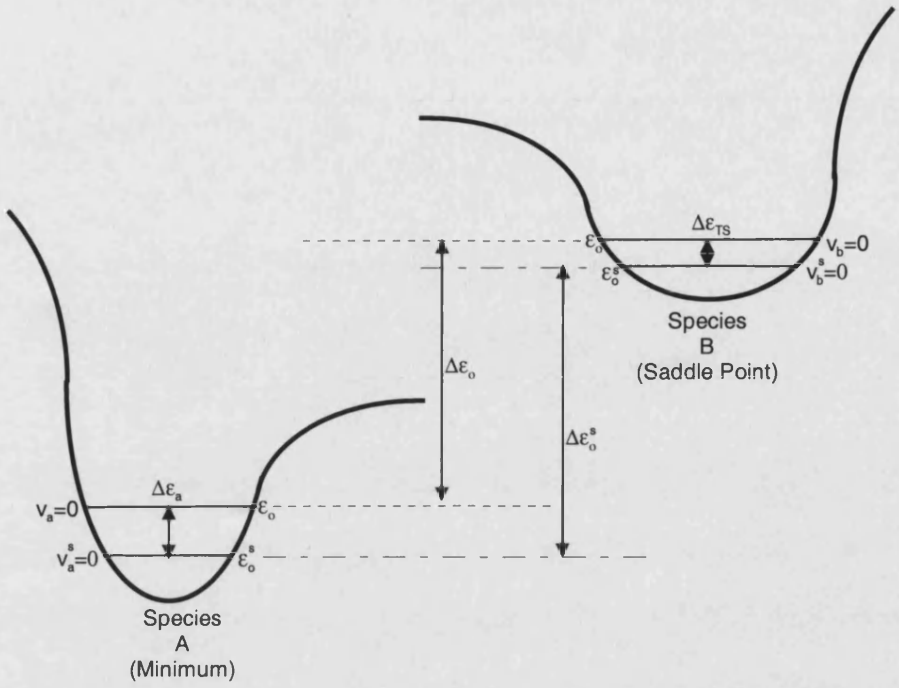


Fig 2.12. Schematic diagram showing the differences in the ZPE along the "reaction co-ordinate" degree of freedom for species A and B, the latter being located at a first-order saddle point on the PES.

An alternative expression for eqn. 2.36 and eqn. 2.37 can be given in terms of the energy gaps between the substituted and unsubstituted molecule for both states under study. For example, the rate constant ratio of eqn 2.37 can be given as:

$$\frac{k}{k^s} = \frac{Q_a}{Q_{TS}} \cdot \frac{Q_{TS}^s}{Q_a^s} \cdot exp-\{(\Delta\varepsilon_o^{TS} - \Delta\varepsilon_o^a)/k_B T\} \qquad (2.38)$$

Here, $\Delta\varepsilon_o^a$ represents the energy gap between the zero vibrational energy level for the substituted and unsubstituted molecule, *A*, and $\Delta\varepsilon_o^{TS}$ is the corresponding difference for the studied TS (figure 2.12). If the energy gap associated with state *A* is larger than that for the TS, then the exponential term is positive and the resulting ratio is more than one. This is known as a *normal isotope effect*; it occurs when the potential of the system becomes looser on moving from reactants to transition-state.

Conversely, if the energy gap associated with the TS is larger, then a negative exponential term is obtained and the rate ratio has a value of less than one. This is

known as an *inverse isotope effect*; it occurs when the potential becomes stiffer on moving from reactants to transition-state.

Primary kinetic isotope effects (1° KIE) involve isotopic substitutions on atoms directly associated with bond making or bond breaking. These effects may have values of about 7.0 for primary deuterium effects, while heavy atom substitutions result in much smaller effects (*ca.* 1.05). *Secondary α-deuterium effects* involve isotopic substitutions on hydrogen atoms bonded to the reacting centre. These atoms are not directly involved with the bond making or bond breaking process and a smaller isotope effect is likely to result compared to primary effects. The rate ratio values obtained for secondary α-deuterium effects tend to be in the range of 0.80-1.25.

More appropriate substitutions as regards the studies presented in this thesis are *secondary β-deuterium effects* (2°-β-KIE). These result from isotopic substitution involving hydrogens on the adjacent site to the reaction centre. Their effect can be given by:

$$ln(k_H / k_D) = cos^2\theta\, ln(k_H / k_D)_{max} + (k_H / k_D)_i \qquad (2.39)$$

The first term on the right hand side of eqn. 2.39 accounts for the hyperconjugation effect obtained from the dihedral alignment between one of the β-hydrons and the developing empty p-orbital at the reacting centre. The empty p-orbital interacts with the anti-bonding orbital of the β-hydron, thus loosening the bond to the hydrogen while double bond character is gained along the *adjacent centre — reactant site* bond. This is discussed further in section 4.1. The maximum hyperconjugative effect $[(k_H/k_D)_{max}]$ is obtained when there is maximum overlap between the empty p-orbital and a β-hydron (i.e. when $\theta=0^{\circ}$).

The second term is a slight electron donating effect associated with the deuterium system (this opposes hyperconjugation). This inductive effect is usually minor compared to the hyperconjugation also taking place.

Finally, a *solvent isotope effect* is obtained when H_2O solvent is replaced with D_2O. This may be due to a primary effect in which solvent participates explicitly in the rate-determining step of a reaction. Additionally, general base catalysis can occur in which a proton is transferred from one water molecule to the other. Furthermore, a secondary effect associated with non-cleaved solvent protons may also exist. The solvent isotope effect therefore, is a complex combination of effects involving many solvent molecules.

2.4.2. Isotope Effects in Enzyme Catalysed Reactions[19]

Michaelis and Menten provided mathematical treatment for the kinetics of an enzyme catalysed reaction involving the conversion of substrate, S, into product, P in the presence of enzyme, E.[20]

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

In this scheme, it is assumed that the substrate first binds to the enzyme to form an enzyme-substrate complex, ES, before forming the product. It is also assumed that the second stage is non-reversible and that only one ES complex is formed during the reaction. The ES concentration can be given in terms of the measurable starting concentration of the enzyme, $[E]_o$, provided that [ES] is constant during the reaction. Using steady state kinetics:

$$\frac{d[ES]}{dt} = k_1([E]_o - [ES])[S] - k_{-1}[ES] - k_2[ES] = 0 \qquad (2.40)$$

From the steady state assumption, an expression can be formulated for [ES]. Using eqn. 2.40:

$$[ES](k_1[S] + k_{-1} + k_2) = k_1[E]_o[S] \qquad (2.41)$$

Rearranging and dividing through by k_1 results in an expression for [ES]:

$$[ES] = \frac{[E]_o[S]}{\frac{k_{-1} + k_2}{k_1} + [S]} \qquad (2.42)$$

The rate constant fraction on the denominator of eqn. 2.42 is known as the Michaelis constant, K_m. It represents a pseudo-equilibrium constant for the formation and destruction of the ES complex over the whole reaction scheme. The overall rate of reaction is then given by $k_2[ES]$:

$$rate = \frac{V_{max}[S]}{K_m + [S]} \tag{2.43}$$

where $V_{max}=k_2[E]_o$ and represents the limiting rate for the enzyme catalysed reaction. If the substrate concentration is in excess, then the value of K_m is insignificant in comparison, and the rate is approximately equal to the limiting rate for the reaction. However, if the substrate is under limiting concentrations, then the value of K_m is more significant than [S] on the denominator. These situations can be summarised as:

$$\text{if} \quad [S] >> K_m \qquad rate \approx V_{max}$$

$$\text{if} \quad [S] << K_m \qquad rate \approx \frac{V_{max}[S]}{K_m}$$

There exists therefore, two conditions under which isotope effects can be evaluated. When [S] is in excess, it is assumed that all the enzyme is complexed and the rate determining step may be a single or a series of chemical changes involving the ES complex. However, under limiting concentrations, the rate will also be partially dependent on substrate binding to the enzyme. Any V_{max}/K_m isotope effect must therefore not only include the chemical step(s) involving the ES complex, but also the EIE for the binding of the substrate to the enzyme as well.

3. SIMULATION METHODS AND APPROACHES

3.1. *Ab initio* Modelling

Ab initio MO calculations were carried out using the GAUSSIAN 92[22] and 94[8] series of programs on local Silicon Graphics R4400 and Hewlett-Packard 720 workstations and on a DEC 8400 computer (Rutherford Appleton Laboratory). The semi-empirical MO calculations were performed using the MOPAC93 program[23] on a local Silicon Graphics R4000 workstation.

Structure and Stability of Hydroxyoxiranone (chapter 4): Geometry optimisation of Hydroxyoxiranone was performed at MP2(full)/6-311G(d,p), with a subsequent single-point energy calculation at the QCISD(T)/6-311G(2df,p) level. In addition to these *ab initio* studies, the structure was also determined using the semi-empirical AM1 level of theory (all MOPAC93 optimisations in this thesis were done with a 0.04 kJ mol^{-1} $Å^{-1}$ gradient tolerance). Minima and saddle points were characterised by determination of their vibrational frequencies.

Intermediates in non-enzymic sialoside hydrolysis (chapter 5): Comparison between the energies of the α-lactone, zwitterion and carbene/carbon dioxide **model I** intermediates (section 5.1) were made at the MP2(fc)/6-31+G(d) level, with diffuse functions on C and O to give a better description of the negative charge on the zwitterion. The geometries for the larger **model II** intermediates (section 5.1) were determined at HF/6-31+G(d), with single-point energy calculations carried out on these structures at MP2(fc)/6-31+G(d). Geometry optimisations were not done at MP2(fc)/6-31+G(d) simply due to CPU time restrictions. Again, characterisation of each stationary point was achieved through the calculation of the harmonic vibrational frequencies. Any counterpoise corrections associated with the carbene/carbon dioxide system were evaluated using the same levels of theory used to determine the energies of the corresponding α-lactone and zwitterion energies. For the *ab initio* solvation studies, single-point isodensity surface polarised

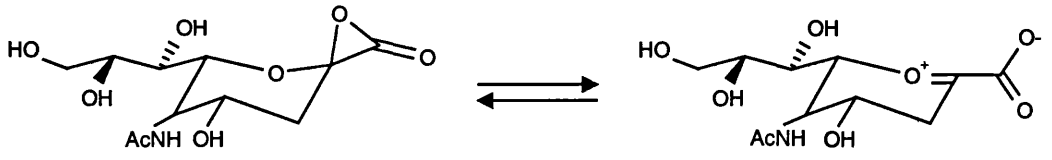
continuum model (IPCM, section 2.2.4) calculations were done at the MP2(fc)/6-31+G(d)//MP2(fc)/6-31+G(d) level. Semi-empirical MO calculations employed the AM1 Hamiltonian and the COSMO (section 2.2.4) method for aqueous solvation. Both techniques used a value of 78.4 for the relative permittivity of water.

3.2. Enzymic and Non-Enzymic QM/MM Simulations

The enzymic and non-enzymic simulations were all done using CHARMM 24b2[11] on the DEC 8400 computer based at the Rutherford-Appleton Laboratory (R.A.L.), Oxford, UK Example CHARMM scripts for both types of simulation are given in appendix 1.

3.2.1. Non-Enzymic QM/MM Simulation Set-up

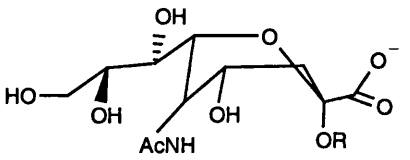
For the non-enzymic simulations, the information from the parameter, topology and co-ordinate files were used to generate the structure of the system. A 15Å stochastic boundary potential was applied to all TIP3P oxygen atoms in the *bulk* sphere (chapter 6); a 10Å potential was used to maintain the smaller *immediate* solvent sphere during the simulated annealing phase. Non-bonded cut-offs were set to 40Å; this effectively included all non-bonded interactions in the QM/MM simulation. Reaction co-ordinates for the α-lactone/zwitterion interconversion (scheme 3.1) were generated by applying an internal co-ordinate constraint to the C_α—O_n bond. Each structure along the reaction profile was subsequently generated from the previous structure by shortening/elongating the C_α—O_n by 0.05Å and allowing the rest of the system to relax around it. Each structure along the profile was minimised using the Adopted-Basis Newton-Raphson (ABNR) technique in the CHARMM program, with re-start structures being saved to disk every 200 cycles. The overall methodology for the solvation simulations is discussed in more detail later (section 6.2).



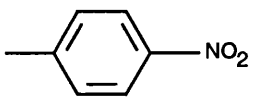
Scheme 3.1

3.2.2. Enzymic QM/MM Simulation Set-up

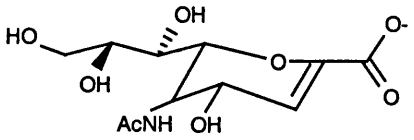
The 2.0Å resolution structure of influenza sialidase complexed with N-acetyl-α-D-neuraminic acid (Neu5ac, **1**) solved by Varghese *et al.*[24] (*protein data bank code:* 2BAT) was used as an initial starting point for the enzyme simulations. One of the principle advantages of using this structure is the appropriate boat conformation that the pyranose ring adopts. The complexes containing inhibitor, 2-deoxy-2,3-dedihydro-N-acetyl neuraminide (Neu5ac2en, **3**), in contrast, have the ring flattened about the C2 centre which is inappropriate for the description of the ES complexes under study. Hydrogen atoms are subsequently added to the crystal structure using the HBUILD command in CHARMM.



R= —H **1**



R= **2**



3

To generate the *p*-nitrophenyl-N-acetyl-α-D-neuraminic acid (Neu5acPNP, **2**) substrate, the PNP group was added to the undocked Neu5ac using QUANTA[25] while keeping the heavy atom positions in the original substrate identical to their positions in the crystal structure. The Neu5acPNP was then docked into the enzyme active site and the resulting ES complex was centred about the substrate.

The next stage in setting-up the system was the addition of extra solvent molecules and the deletion of unwanted atoms from the system. A 17Å water sphere was placed over the ES system and any solvent molecule whose oxygen atom was within 2.5Å of any heavy atoms in the ES complex was deleted. In addition, enzyme residues containing *all* atoms beyond 15Å of the substrate were also deleted. While the removal of overlapping water molecules was carried out to keep the non-bonded

atom repulsions to a minimum, the deletion of enzyme residues was done to reduce CPU time demands.

Next, constraints and boundaries were applied. The deletion of the outer protein residues results in the destruction of the secondary structure around the periphery of the remaining enzyme. To maintain the structural integrity of the protein at these positions, a harmonic force constraint was applied to all enzyme and crystal water atoms located more than 13Å away from the origin. This constraint, however, allows limited mobility of these residues during the optimisation procedure. As well as the harmonic constraint applied to the outer protein and crystal water atoms, a 19Å stochastic boundary was applied to all the water molecules within the system. The resulting cut-down model is defined as *es0*.

The ES starting structures (where the substrate refers to Neu5ac or Neu5acPNP) were generated as follows (figure 3.1). From *es0*, all water molecules were allowed to relax around the remaining fixed system (giving structure *es1*). From *es1*, the protein hydrogen atoms were released and optimised with the water molecules (*es2*). Using this structure, the protein side chains were released and the unconstrained part of the system was optimised around the fixed substrate and protein backbone (*es3*). The protein backbone was then released and allowed to optimise with the rest of the protein and the solvent molecules (*es4*). Structures *es0*, *es1*, *es2*, *es3* and *es4* were described with an MM only potential and structures *es1-es4* were optimised to an RMS gradient tolerance of 0.04 kJ mol^{-1} $Å^{-1}$ (0.01 kcal mol^{-1} $Å^{-1}$). Similar gradual optimisation procedures has been employed by others when dealing with proteins.[30,31]

The next stage of this process involved treating the substrate quantum mechanically and allowing the entire QM/MM system to optimise until the RMS gradient tolerance of 0.04 kJ mol^{-1} $Å^{-1}$ was satisfied (giving *es5*). The energy of *es5* compares favourably to a system subjected only to a two-step optimisation procedure (*opt2s*). Such a two-step process involved MM optimisation of protein and solvent only, followed by QM/MM optimisation of the ES complex (in which the QM substrate was unconstrained). The energy difference, ΔE_{tot}, between *es5* and *opt2s* was some

200 kJ mol^{-1} showing that a gradual optimisation procedure results in a system with a much lower energy.

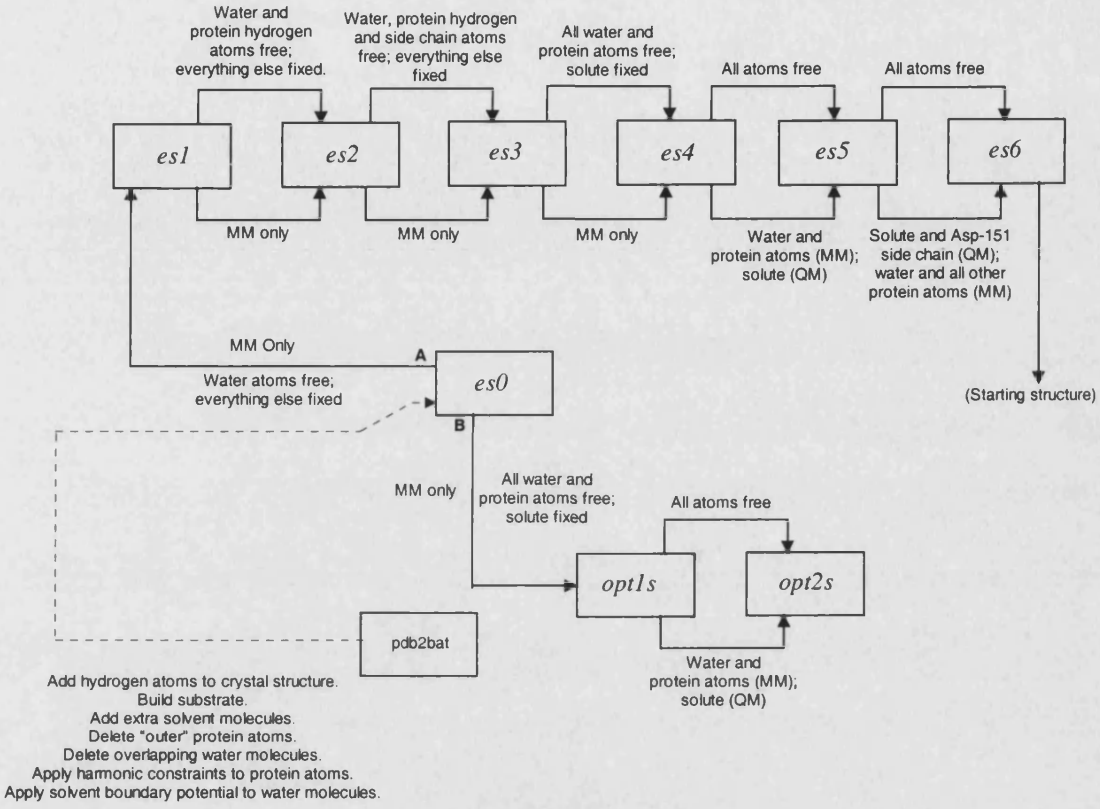


Fig 3.1. Two possible methods of optimising the ES complex. Path **A** represents a more complex gradual optimisation route. Path **B** represents a simpler two step process. "pdb2bat" is the original crystal structure.

The final stage in generating a starting ES structure is to treat the Asp-151 side chain quantum mechanically (see chapter 8 for mechanistic postulations); this requires the use of a QM link atom. In this thesis, it was decided to fix the position of the QM link atom throughout the simulation; such a methodology has also been used by Hart *et al.* in their QM/MM study of protein tyrosine phosphatase.[26] Because the QM link atom is fixed, a reasonable position for it must be found in the ES structure before proceeding with the simulation. To determine its optimal position, the QM link atom was translated from the QM C_β atom to the MM C_α atom over 20 steps, with the energy of the system being calculated at each stage. The potential produced (figure 3.2) has a minimum when the QM link atom was about 1.16Å away from the C_β carbon. The QM link atom was subsequently placed at this distance on the *es5* structure and fixed in space. The other atoms were subsequently allowed to optimise

around it until a gradient tolerance of 0.02 kJ mol^{-1} $Å^{-1}$ was satisfied. The resulting structure, *es6*, was then used as the initial structure for the simulation. It should also be noted that the Asp-151 carboxylate group is protonated resulting in a QM system with an overall charge of -1. The protonated residue is referred to as Asp(H)-151 in this thesis.

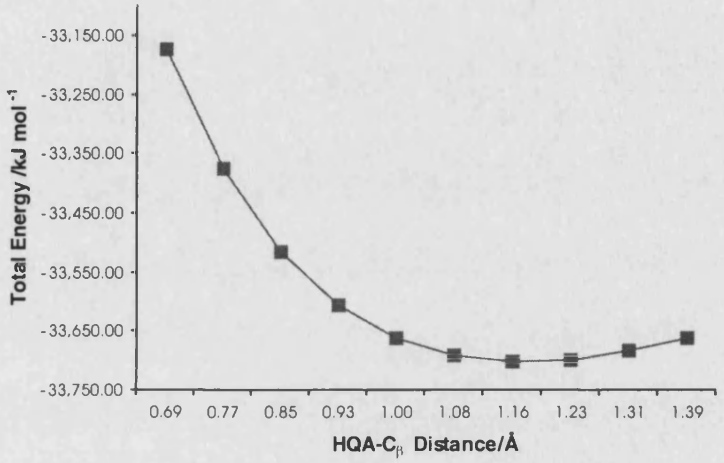


Fig 3.2. The potential obtained when a QM link atom (HQA) is moved from the QM described C_β atom to the MM described C_α atom. The QM link atom is translated along the C_α—C_β bond axis in Asp(H)-151.

3.3. General Re-programmable Algebraic Chemistry Engine (GRACE)

While CHARMM provides the basis for MM or QM/MM minimisation and dynamics simulations, the program lacks some of the functionality required for modelling chemical reactivity. CHARMM does not have the ability to locate and characterise transition structures or perform IRC calculations. Turner has therefore developed GRACE,[27] a program that includes such functionality; its principle advantage is to treat systems consisting of many thousands of atoms. GRACE also provides the functionality to combine several different QM codes (for example, GAMESS and CADPAC) with the CHARMM force field and thus gives the opportunity to use an *ab initio*/CHARMM potential for describing the chemical species of interest. However, while GRACE is a powerful tool for modelling chemical reactivity in large systems, its disadvantage is the user unfriendliness of the input scripts and the lack of documentation available.

The GRACE approach to modelling large species involves dividing the system into two sections. Firstly, a CORE defines a sub-set of atoms found in and around the reaction centre, while the ENVIRONMENT includes the rest of the system. Such a procedure is employed within GRACE because the manipulation of the hessian matrix is computationally intractable for large systems such as proteins. The hessian is therefore only evaluated for the most essential atoms within the system (the CORE). The CORE and ENVIRONMENT partition need not correspond to the quantum mechanics and molecular mechanics partition in a QM/MM system. The CORE and ENVIRONMENT are therefore able to incorporate both QM and MM described atoms.

3.3.1. Transition State Refinement

The procedure of TS refinement calculates the first and second derivatives of the CORE while trying to keep the gradient of the ENVIRONMENT below a certain gradient tolerance (figure 3.3). After optimising the ENVIRONMENT around the fixed CORE and generation of the initial hessian matrix, the process of TS refinement begins. After each step taken by the partial rational function operator (P-RFO) optimiser, the CORE atoms are fixed and the ENVIRONMENT is allowed to optimise around them. The P-RFO optimiser follows a specified mode in the hessian and maximises the energy in this direction while minimising in all the others. The hessian matrix is then up-dated (generation of a new hessian after each optimisation stage would be time consuming)

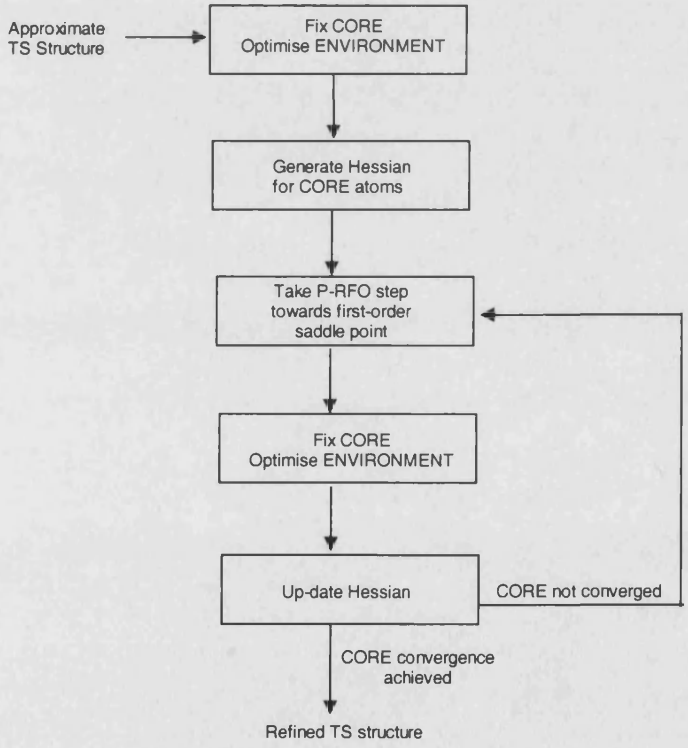


Fig 3.3. Basic procedure for TS refinement in GRACE

and the process is repeated. A successful refinement leads to a first-order saddle point with respect to the CORE atoms, surrounded by a minimised ENVIRONMENT. For all TS refinements carried out in this thesis, the exit criterion of 0.02kJ mol^{-1} $Å^{-1}$ was used for the ENVIRONMENT RMS gradient. The maximum and RMS gradient tolerances for the CORE were set to 1.26 kJ mol^{-1} $Å^{-1}$ and 0.42 kJ mol^{-1} $Å^{-1}$ respectively, in line with the default exit criteria for optimisations in programs such as GAUSSIAN.[8]

3.3.2. The GRACE Intrinsic Reaction Co-ordinate Approach

In GRACE, the hessian for the CORE atoms in the refined TS structure is first evaluated and a check is done to verify the presence of one negative eigenvalue. If true, the IRC is followed towards reactants or products. The RMS gradient tolerance is set to 0.42 kJ mol^{-1}, but the IRC is also terminated when the energy of the system increases.

The approach taken in GRACE is as follows. At each step taken along the IRC path, the ENVIRONMENT is minimised around the fixed CORE before the next step is taken. For IRCs calculated in this thesis, the ENVIRONMENT gradient tolerance was set to 0.02 kJ mol^{-1} $Å^{-1}$. This approach means that while a genuine IRC is followed for the CORE atoms, the ENVIRONMENT does not follow the steepest descents profile. The ENVIRONMENT can therefore be subjected to an irreversible change resulting in an unrelated structure to that obtained in the proceeding step. This could lead to larger displacements in the ENVIRONMENT atoms than perhaps would be obtained if they were explicitly included in the IRC approach. Such a change in geometry may result in a corresponding increase in the total energy of the system and thus might result in a premature exit of IRC calculation. Therefore all IRC reactant and product structures determined in this thesis were subject to further optimisation in CHARMM using the ABNR optimiser with an RMS gradient tolerance of 0.02 kJ mol^{-1} $Å^{-1}$.

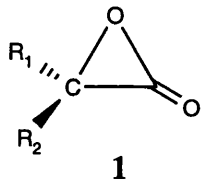
3.3.3. Calculation of Kinetic Isotope Effects

The mass-weighted hessian matrix calculated for the CORE atoms often contains rotation and translation contamination: the six eigenvalues associated with these motions will not exactly be zero. This contamination may also affect the internal co-ordinate vibrations in the system resulting in errors for the calculated vibrational modes and subsequent calculated isotope effects. GRACE has also been used to remove the rototranslational contamination from the hessian (appendix 2c).[28,29] The cleaned hessian for both species under study is subsequently used to form part of the input script for the CAMISO[21] program (appendix 3).

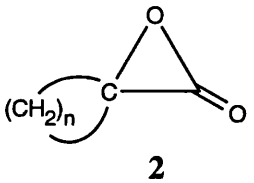
4. STRUCTURE AND STABILITY OF HYDROXYOXIRANONE

4.1. Introduction

α-Lactones (**1**,**2**) are the smallest cyclic esters consisting of a three-membered ring and an exocyclic double bond. Their inherent instability has been attributed to the large amount of *Baeyer strain*[32] associated with the ring, which is enhanced by the presence of the carbonyl double bond.[33]



1



2

Table 4.1. Various substituted α-lactones with assigned carbonyl stretching IR frequencies.

Structure	R_1	R_2	ν(C=O)/cm^{-1}	Structure	n	ν(C=O)/cm^{-1}
1a[e]	H	H	1967	**2a**[f]	2	1935
1b[f]	CH_3	CH_3	1900	**2b**[f]	3	1917[c]
1c[f]	C_4H_9	C_4H_9	1893	**2c**[f]	4	1901
1d[g]	Ph	Ph	1890[a]			
1e[h]	CF_3	CF_3	1980			
1f[i]	$CF_3C(C_2F_5)_2$	$CF_3CF(C_2H_5)$	1990[b]			
1g[j]	Ph	Cl	1920			
1h[j]	PNP[d]	Cl	1928			

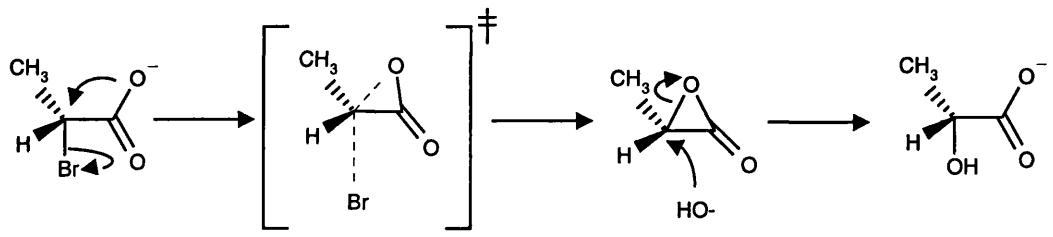
[a]Three frequencies obtained, the others reported as 1878cm^{-1} and 1910cm^{-1}; also, [b]1945cm^{-1} and [c]1895cm^{-1}.[d]*p*-nitrophenyl group. [e]ref 34. [f]ref 35. [g]ref 36. [h]ref 37. [I]ref 38. [j]ref 39.

The formation of α-lactones has been achieved by reacting carbenes with carbon dioxide under matrix-isolation conditions;[34,36,39] they have also been produced by the photo-decarboxylation of 1,2-dioxalane-3,5-diones.[35,40] It has been suggested that the relative stability of α-lactones depends on the stabilising effect of α-substituents upon the adjacent electrophilic carbon atom.[40] Thus, electron-donating species will demote the formation of the ring system in favour for the 1,3-dipolar species,[41]

whereas electron-withdrawing substituents would favour the α-lactone.[37,38] In addition to the thermodynamic stabilisation through electronic effects, a kinetic stabilisation can be achieved through steric hindrance by the inclusion of bulky side groups at the α-centre.[37,41] Formation of α-lactone derivatives containing electron-withdrawing fluoro-alkyl groups have been reported: bis-(trifluoromethyl)oxiranone (**1e**) has been reported to be stable "for many days at 253K (–20°C)".[37] The more stable perfluoro-[(1-ethyl-1-methylpropyl)(1-methylpropyl)]oxiranone (**1f**), which contains bulky, substituted fluoro-alkyl groups, has been reported "to be stable at 263K (–10°C) for up to one year".[42] In contrast, bis-(*tert*-butyl)oxiranone polymerised at 253K (–20°C), affording the polyester as a white precipitate.[41] It is evident that while *tert*-butyl groups provide a degree of kinetic stabilisation, they still allow nucleophilic attack to occur at the electrophilic centre. However, it has been shown for the fluoro-alkyl derivatives that nucleophilic attack takes place preferentially at the carbonyl carbon.[37,42]

α-Lactones have been observed, both directly and indirectly by IR,[34-39] mass spectrometry[43] and NMR[41] techniques. Infrared spectroscopy points evidence towards the formation of α-lactones through a carbonyl stretch. This is located at higher frequencies (≈1900cm^{-1} and above) compared to polyesters (≈1850cm^{-1}), ketones and carboxylic acids (≈1750cm^{-1}). Furthermore, it has been observed that such stretching vibrations are shifted to lower frequencies as the size of the α-lactone ring is increased[44,45]. The carbonyl stretching frequencies assigned to specific α-lactones are given in table 4.1, which include frequencies for spiro-α-lactones that have been formed from the photochemical decomposition of 1,2-dioxalane-3,5-diones at 77K (–196°C).[35]

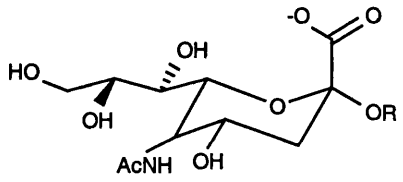
α-Lactones were first postulated as intermediates[46] in the nucleophilic substitution of α-bromopropionate anions,[47] resulting from the neighbouring group participation (NGP) of the adjacent carboxylate group.[48] A mechanism involving such NGP has been suggested to facilitate the removal of the bromide anion from the α-centre and account for the retention of configuration in the substituted product (Scheme 4.1).



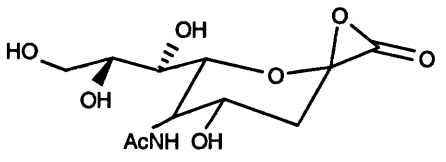
Scheme 4.1.

α-Lactones have also been postulated as transient intermediates in thermal[41,43,49-51] and photochemical[35] degradation; free radical reactions;[52-54] auto-oxidation;[55,56] ozonolysis;[41] ketene oxidation,[57,58] and in the Pig Liver Esterase catalysed Meinwald re-arrangement of epoxy di-esters.[59,60] NGP by carboxylate groups have also been reported to effect the reactivity of carbenes.[61] α-Lactones undergo polymerisation, decarbonylation and nucleophilic attack to yield polyesters, ketones and α-substituted carboxylic acids respectively. The formation and destruction of α-lactones are summarised in scheme 4.2.

The author's interest in α-lactone chemistry originated from experimental kinetic studies of non-enzymic hydrolysis of *p*-nitrophenyl-N-acetyl-α-D-neuraminic acid (**3**, R = PNP), in which the three membered ring species was postulated as a possible intermediate (**4**).[1]



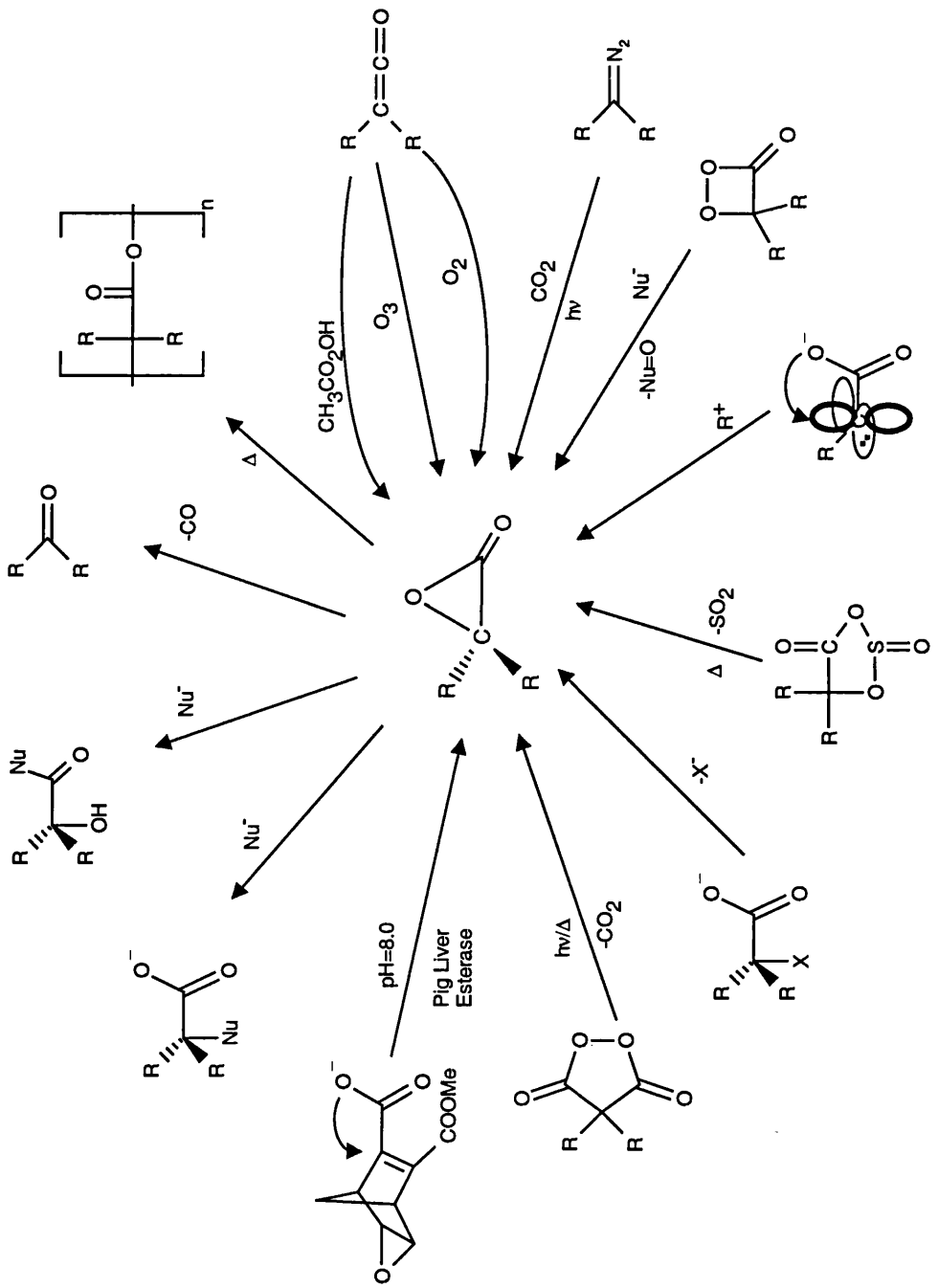
3



4

Evidence for the NGP in the sialoside system was based upon:

1). The inverse solvent isotope effect observed on the departure of the *p*-nitrophenyl leaving group from the sialoside anion.


Δ
-CO
Nu⁻
Nu⁻
pH=8.0
Pig Liver Esterase
hv/Δ
-CO2
-X⁻
Δ
-SO2
R+
-Nu=O
Nu⁻
hv
CO2
O2
O3
CH3CO2OH
COOMe

2). The larger values obtained for the maximum β-deuterium kinetic isotope effect (KIE) at pH 1.0 compared to those obtained at higher levels of pH.

The inverse solvent isotope effect of 0.86 provided circumstantial evidence for the existence of NGP: such an isotope effect requires the destruction of solute-solvent hydrogen bonds in the course of the reaction. For NGP to occur, the carboxylate group must interact with the developing positive charge on the adjacent centre with a concomitant release of hydrogen bond donors into bulk solvent.

The hyperconjugative component of the secondary β-deuterium KIEs gives an indication of the orbital eclipsing between the C-D bond and the increasingly empty p-orbital at the reaction site (Fig 4.1). The maximum effect, $(k_H/k_D)_{max}$, occurs when the C-D bond and the empty p-orbital are fully eclipsed (θ=0). The natural logarithm of $(k_H/k_D)_{max}$ at pH 1.0 was found to be twice as large than under the conditions in which the anion was the predominant species. Thus, the non-labelled reaction was increasingly favoured at lower pH, suggesting a greater amount of hyperconjugation present. This effect was observed to be less important for the neuraminide anion and thus suggested a reduced p-orbital eclipsing with the C-D bond. This can occur if the C2 carbon is deviated away from a trigonal planer structure, which reduces the hyperconjugative effect. This observation points towards the formation of a highly strained α-lactone intermediate (scheme 4.3) through a transition state that involves a degree of NGP.[1]

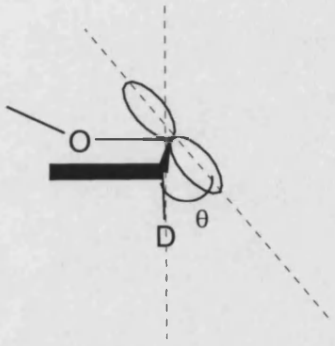
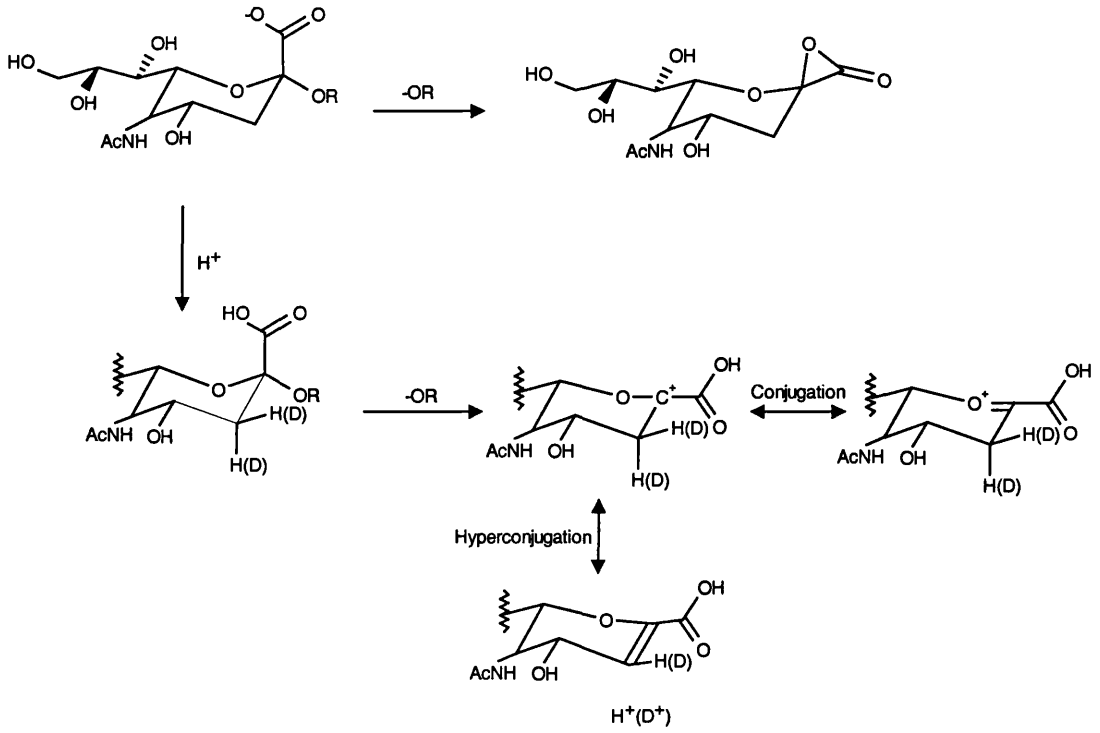
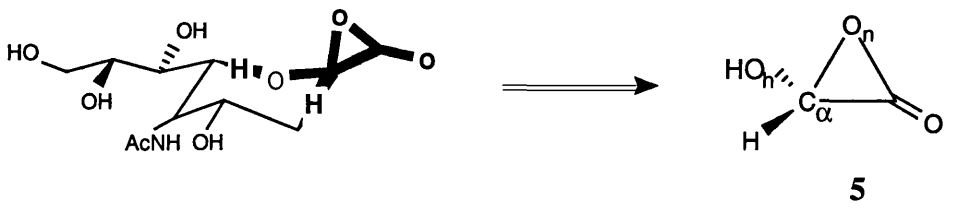


Fig 4.1. Definition of angle, θ.



Scheme 4.3

This chapter reports on an *ab initio* investigation into the structure and stability of the fully formed α-lactone, Hydroxyoxiranone, **5**, *in vacuo*. Figure 4.2 shows its relationship to the full sialoside molecule. In chapter 5, Hydroxyoxiranone is used as a model for the sialoside α-lactone intermediate, but for the work presented here, it is treated as a molecule in its own right.

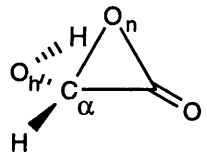


5

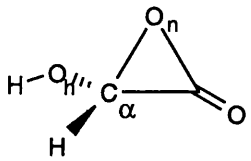
Fig 4.2. Relation of hydroxyoxiranone to the full sialoside α-lactone intermediate

After a brief look at the possible rotamers of **5**, an investigation into the standard heat of formation is given. This is first calculated using the methodology devised by Pople *et al.*[62] and is then compared with the value determined through an isodesmic relationship. A discussion then follows on the relative stability and structural aspects of hydroxyoxiranone, which is then followed by a conclusion.

4.2. Rotamers of Hydroxyoxiranone



Anti



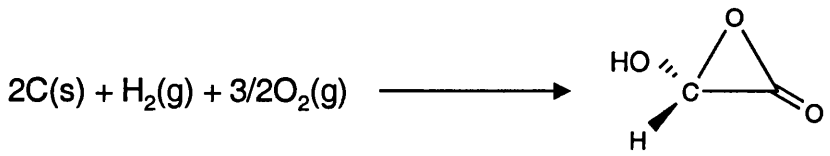
Gauche, **5g**

Hydroxyoxiranone ($C_2H_2O_3$) has one rotational degree of freedom: the rotation about the C-OH bond, which was assumed to have a 3-fold axis of rotation. At MP2/6-31+G(d), only two conformers of $C_2H_2O_3$ were found. The gauche conformer had an H-C_α-O_α-H dihedral angle of 36° and was 5.7 kJ mol^{-1} higher in energy. The anti rotamer had a dihedral angle of 170°, with the H-C_α and O_α-H bonds approximately antiperiplaner to one another (table 4.3). A rotamer with a dihedral angle of –70° was found to lead to **5g** when optimised. At the semi-empirical AM1 level of theory, only the anti conformer was obtained *in vacuo*.

The more stable anti conformer was used to determine the standard heat of formation of hydroxyoxiranone.

4.3. Calculation of the Standard Heat of Formation from Atomisation Energies

At a specified temperature, the standard heat of formation, ΔH_f, for a compound is given relative to the ΔH_f of its constituent elements (in their reference states) at 1 atm. Thus $\Delta H_{f,298}$ for hydroxyoxiranone is considered to be the energy change for the following formal reaction:



(4.1)

This can also be defined using the atomisation enthalpies (at a certain temperature) for both the compound and its constituent elements such that:

$$\Delta H^\circ_{f,298} = \Delta H_{atm(hydroxyoxiranone)} - \Delta H_{atm\ (elements)} \qquad (4.2)$$

This is schematically shown in figure 4.3.

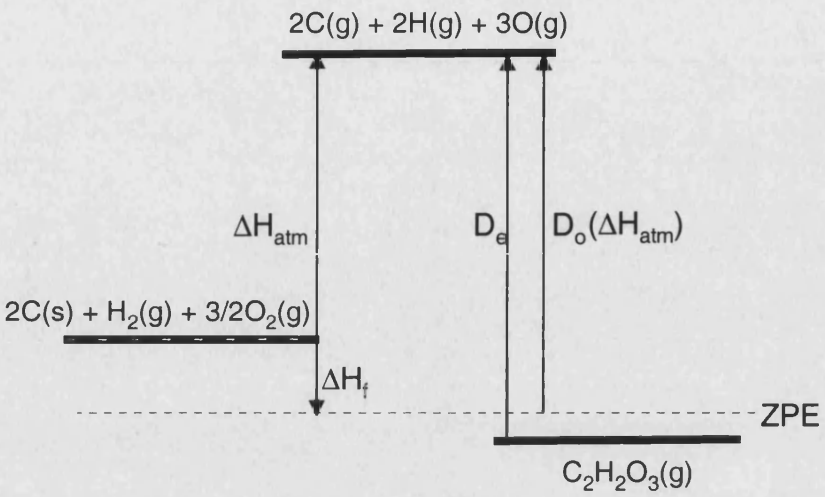


Fig 4.3. Relationship of ΔH_f with the atomisation enthalpies (ΔH_{atm}) of hydroxyoxiranone and corresponding elements.

4.3.1. Calculation of Atomisation Energies

The QCISD(T)(full)/6-311G(2df,p) energies for hydroxyoxiranone, hydrogen and the elements are given in Table 4.2.

Table 4.2. QCISD(T)(Full)/6-311G(2df,p) energies for hydroxyoxiranone and its elements (in hartrees).

$C_2H_2O_3$	H_2	H	C	O
-302.685572	-1.168323	-0.499810	-37.796423	-74.986014

The spectroscopic dissociation energy, D_e, can be calculated using:

$$C_2H_2O_3\ (g) \xrightarrow{\Delta E_1} 2C(g) + 2H(g) + 3O(g) \qquad (4.3)$$

$$\Delta E_1 = 2(-37.796423) + 2(-0.499810) + 3(-74.986014) - (-302.685572)$$

$$\Delta E_1 = +1.1350656\ E_h$$

$$(1\ hartree = 2625\ kJ\ mol^{-1})$$

However, equation 4.3, is non-isogyric: in the ground state, the total number of unpaired electrons in the elements is 12. The inclusion of six molecules of H_2 into equation 4.3 gives the isogyric equation:

$$C_2H_2O_3\,(g) + 10H(g) \xrightarrow{\Delta E_2} 2C(g) + 3O(g) + 6H_2(g) \qquad (4.4)$$

$$\Delta E_2 = \Delta E_1 + 6(-1.168323) - 10(-0.499810)$$

$$\Delta E_2 = +0.122845\ E_h$$

However, ΔE_2, also includes the formation of six molecules of hydrogen. The exact dissociation energy of H_2 at 0K (0.17447 E_h)[63] is used in order to account for this discrepancy. The new value of ΔE_2 becomes:

$$\Delta E_2' = \Delta E_2 - 6(-0.17447) = +1.1696652\ E_h\ (3070.06\ kJ\ mol^{-1})$$

This is the spectroscopic dissociation energy of hydroxyoxiranone. The atomisation energy (D_0) is computed by subtracting the zero-point vibrational energy (ZPE, determined at HF/6-31G(d,p) and scaled by 0.89, table 4.4). Thus:

$$\mathbf{\Delta H_{atm(hydroxyoxiranone)} = 2959\ kJ\ mol^{-1}}$$

The atomisation energies of the elements can be calculated using the published $\Delta H_{f,0}$ for the gaseous atoms (C, 711; H, 216 and O, 247 kJ mol^{-1}):[64]

$$2C(s) + H_2(g) + 3/2O_2 \xrightarrow{\Delta E_3} 2C(g) + 2H(g) + 3O(g) \qquad (4.5)$$

$$\Delta E_3 = 2(711) + 2(216) + 3(247)$$

$$\Delta E_3 = 2595\ kJ\ mol^{-1}$$

Therefore,

$$\mathbf{\Delta H_{atm\ (Elements)} = 2595\ kJ\ mol^{-1}}$$

This leads to

$$\mathbf{\Delta H_{f,0\ (hydroxyoxiranone)} = -364\ kJ\ mol^{-1}}$$

4.3.2. Addition of Enthalpy Corrections to Give $\Delta H_{f,298}$ for Hydroxyoxiranone

It is usual to report heats of formation at 298K (25°C). Taking the calculated ZPE and internal energy of hydroxyoxiranone, along with the published heat capacities of the elements in their standard states,[65] $\Delta H_{f,298}$ for $C_2H_2O_3$ can be determined.

It has been shown that for any system at mechanical equilibrium:

$$H = U + PV \tag{4.6}$$

Where PV = RT for one mole of perfect gas; H is the enthalpy and U is the internal energy. It can also be shown that at constant pressure:

$$\Delta H = \int C_p dT \tag{4.7}$$

where C_p is the heat capacity of a substance at constant pressure and T is the temperature. Further, the enthalpies at two different temperatures can be related by:

$$H(T_2) = H(T_1) + \Delta H \tag{4.8}$$

Substituting H and ΔH with expressions in eqn 4.6 and eqn 4.7 respectively:

$$U_{T2} + RT_2 = U_{T1} + RT_1 + \int_{T_1}^{T_2} C_p dT \tag{4.9}$$

Because T_1 is 0K, the enthalpy correction can be given in terms of the internal energy differential (the thermal correction, E_{th}) and the RT_2 work expansion term:

$$(U_{T2} - U_{T1}) + RT_2 = \int_{T_1}^{T_2} C_p dT \tag{4.10}$$

U_{T2} can be obtained from the thermochemistry analysis of hydroxyoxiranone and U_{T1} is the ZPE; T_2 is 298K and R is the molar gas constant. Thus, using eqn. 4.10 and the U_{298} of 137 kJ mol^{-1} for hydroxyoxiranone:

Enthalpy Correction for Hydroxyoxiranone = 13.8 kJ mol^{-1}

Using the constant-pressure heat capacities[65] for C, H and O, the enthalpy correction for one mole of each element in their standard states is 1.050, 8.468 and 8.682 kJ mol^{-1} respectively. Allowing for elemental stoichiometry:

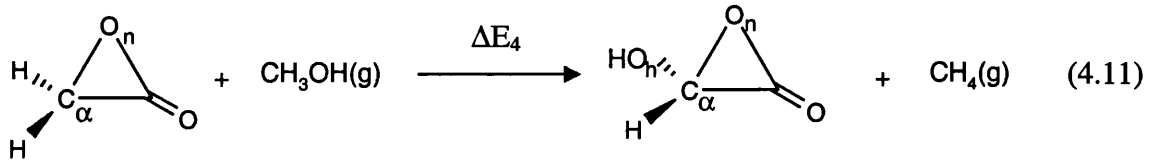
Enthalpy Correction for Elements = 23.6 kJ mol^{-1}

Thus, taking these enthalpy corrections into account:

$$\Delta H_{f,298\ (hydroxyoxiranone)} = -374\ kJ\ mol^{-1}$$

4.4. Calculation of $\Delta H_{f,298}$ Using an Isodesmic Relationship

The standard heat of formation of $C_2H_2O_3$ can be also obtained from the following isodesmic relationship:



(4.11)

The energy change for this reaction was found to be –65.0 kJ mol^{-1} at the QCISD(T)/6-311G(2df,p)//MP2/6-311G(d,p) level of theory (Table 4.4). The enthalpy change for the reaction at 298K was determined using eqn 4.10 to calculate the thermal corrections for all reactants and products. This gave a $\Delta H_{r,298}$ of –62.1 kJ mol^{-1}.

The calculated $\Delta H_{f,298}$ for oxiranone has been reported[33] as -190 ±10 kJ mol^{-1}. Furthermore, the experimental $\Delta H_{f,298}$ for gaseous methanol and methane has been determined as -201.7 kJ mol^{-1} and -74.5 kJ mol^{-1} respectively.[64] Using these values, along with the corrected enthalpy of reaction, the standard heat of formation for hydroxyoxiranone is calculated as:

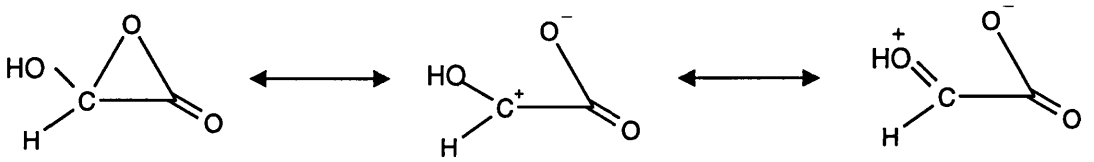
$$\Delta H_{f,298}(C_2H_2O_3) = \Delta H_{r,298} + \Delta H_{f,298}(C_2H_2O_2) + \Delta H_{f,298}(CH_3OH) - \Delta H_{f,298}(CH_4)$$

$$\Delta H_{f,298}(\mathbf{C_2H_2O_3}) = (-62.1) + (-190\pm10) + (-201.7) - (-74.5) = \mathbf{-379kJ\ mol^{-1}}$$

This value compares well with the –374 kJ mol^{-1} obtained using the method in section 4.3. An estimate of the **standard heat of formation of hydroxyoxiranone at 298K** can be considered as the average of the values, i.e. $\Delta H_{f,298}$= **-377 ± 10 kJ mol^{-1}**. The error has been estimated from the combination of errors associated with the average of the two calculated heats of formation for $C_2H_2O_3$ (*ca.* 2 kJ mol^{-1}) and the reported error associated with the Pople method (*ca.* 8 kJ mol^{-1}).[62]

4.5. Discussion

The exothermic change for the isodesmic relationship (eqn 4.11) provides a measure of the reduction in ring strain in $C_2H_2O_3$ as compared with oxiranone. At MP2/6-31+G(d), introduction of the hydroxyl substituent at the α-position affords an increase in the C_α—O_n by 0.035Å as compared with the unsubstituted α-lactone. This is accompanied by increasing planarity about the C_α atom: ΣC_α for hydroxyoxiranone is closer to 360°. Further, the C_α—O_h distance in $C_2H_2O_3$ is some 0.08Å shorter than a standard C—O single bond. These observations suggest a structure that is resonance stabilised as shown in scheme 4.4. In molecular orbital terms, there is a degree of n_π electron donation from the O_h atom to the adjacent p_π orbital on the C_α atom.



Scheme 4.4

Many of these structural changes are increasingly apparent when larger basis sets and electron correlation are used to describe the system. The exothermic change for eqn. 4.11 becomes more pronounced: at MP2/6-311G(d,p), ΔE_4 is some 16.6 kJ mol^{-1} more exothermic than at HF/6-31G(d,p). However, single-point QCISD(T)

calculations on the MP2/6-311G(d,p) hydroxyoxiranone structure give no further stabilisation over oxiranone.

The hydroxyl substituent relieves a degree of strain in the hydroxyoxiranone ring by promoting the elongation of the C_{α}-O_n bond. The ring strain energy of oxiranone has been estimated by Rodriquez and Williams[33] to be 169 kJ mol^{-1}. From this current study, the strain energy for hydroxyoxiranone can be evaluated as 104 kJ mol^{-1}, which compares well with the ring strain energies of oxirane (114 kJ mol^{-1}) and cyclopropane (115 kJ mol^{-1}). On this result alone, there is no reason why hydroxyoxiranone could not be produced by many of the reactions given in scheme 4.2. It is probable that $C_2H_2O_3$ would only have a relatively short lifetime due to its enhanced 1,3-dipolar character.

Other properties of α-lactones arouse interest. IR spectroscopy can give information on the carbonyl stretching frequency, of which, reported values for several α-lactones are listed in table 4.1. The frequency of the carbonyl absorption is known to be inversely proportional to ring size in lactones.[44,45] This suggests that it is possible to formulate a series of α-lactones with respect to their carbonyl stretching frequency. Such a series could reveal qualitatively, the effect of the R substituent on ring closure over the development of 1,3-dipolar character in the system.

Electron-withdrawing substituents on the α-carbon would favour ring closure over the development of zwitterionic character, with the reverse being true for electron-donating substituents. The more fully formed α-lactone structure would be expected to have a carbonyl stretch shifted to higher frequency as compared to more 1,3-dipolar systems.

A series allowing evaluation of the individual substituent contributions to ring strain would require more data than is currently available. Thus, at present, a series could only be given in terms of the combined effect of both substituents at the α-centre. Further complications arise from the differing conditions under which the IR spectra were taken. α-Lactone, **1e**, for example, has the carbonyl absorption at 1980 cm^{-1} in the vapour phase. However, when in the condensed phase or in CCl_4 solution, the

carbonyl absorptions for **1e** occur at 1975 cm^{-1} and 1970 cm^{-1} respectively.[37] Furthermore, solvent effects may be important, not only in influencing the IR absorption positions, but also on the structure of the α-lactone itself.

Therefore, these complications suggest that α-lactone structures can only be strictly compared if they are:

1). Characterised in the vapour phase (**1e**) or,
2). Prepared and characterised under matrix isolation conditions (**1a**, **1d**, **1g** and **1h**).

Assuming the carbonyl IR frequency for a given α-lactone under matrix isolation conditions is similar to that obtained in the vapour phase, then:

1d < 1g < 1h < 1a < 1e

From this mini-series, it is evident that electron-withdrawing groups result in a shift of the carbonyl stretch to higher frequencies. Indeed, this is also the case if differences in the preparative conditions are overlooked, giving the series:

1d < 1c < 1b < 1g < 1h < 1a < 1e < 1f

It is also note-worthy that for the spiro-fused systems, the frequency of the carbonyl stretch is inversely proportional to the size of the ring attached onto the α-lactone. Thus, a larger value for n in **2** gives the carbonyl absorption at lower frequencies.

Table 4.5 gives the unscaled calculated frequencies for the carbonyl stretch at HF/6-31G(d,p) and MP2/6-31+G(d) in hydroxyoxiranone. Also included are the calculated frequencies for oxiranone and methyl ethanoate.

Table 4.5. Unscaled calculated IR frequencies for the carbonyl stretch absorption of hydroxyoxiranone (anti and gauche conformers), oxiranone and methyl ethanoate. Experimental value for the ester is the average of the reported values in the vapour phase (refs 66 and 67). The experimental value for oxiranone refers to the IR frequency obtained under matrix isolation conditions (ref 34).

Compound	HF/6-31G(d,p)/ cm^{-1}	MP2/6-31+G(d)/ cm^{-1}	Expt/ cm^{-1}
Hydroxyoxiranone			
Anti	2212	1959	
Gauche	2229	1983	
Oxiranone	2217	1984	1967
Methyl Ethanoate	2018	1792	1775

Comparing the calculated carbonyl frequency for methyl ethanoate at HF/6-31G(d,p) with the experimental value,[66,67] it is evident the calculation has over estimated the frequency by some 14%. Using the resulting scale factor of 0.8796 at this level of theory, the estimated carbonyl stretching frequency for oxiranone is 1950 cm^{-1}. At MP2/6-31+G(d), the methyl ethanoate frequency is overestimated by 1%. Using a scale factor of 0.9905 affords a calculated stretching frequency for oxiranone of 1965 cm^{-1}.

The scaled frequencies for hydroxyoxiranone at MP2/6-31+G(d) are 1940 cm^{-1} and 1964 cm^{-1} for the anti- and gauche conformers respectively. From these values, it is likely that the anti conformer will be more strained than **1g** and **1h**, but less strained than oxiranone. Interestingly, at MP2/6-31+G(d), the carbonyl absorption for the gauche conformer is similar to that obtained for oxiranone. This emphasises that a series based on IR frequencies of carbonyl stretches could only be qualitative. However, the calculated stretches also suggest the anti conformer to be less strained over the gauche; as can been seen by table 4.3, the C_α—O_n distance is slightly smaller as compared with the anti conformer.

Coe *et al.* reported two frequencies for the carbonyl absorption in liquid **1f**.[38] Similarly, Sander reported three frequencies for **1d** in CO_2 matrices.[36] Such effects have been attributed to different substituent conformations. In this present study, it

may be predicted that two carbonyl absorption bands are obtained for hydroxyoxiranone due to a change in the hydroxyl substituent conformation.

4.6. Conclusions

α-Lactones have been isolated as separate species and identified as possible intermediates in several reactions. This study has calculated the heat of formation of the anti-conformer of hydroxyoxiranone to be -377 ± 10 kJ mol^{-1}. The structure of $C_2H_2O_3$ was found to be less strained than oxiranone due to possible resonance stabilisation involving the hydroxyl oxygen. It was found that hydroxyoxiranone has two IR bands assigned to the carbonyl stretch at MP2/6-31+G(d), which when scaled, have values of 1940 cm^{-1} and 1965 cm^{-1}. These values refer to the anti- and gauche conformers of hydroxyoxiranone obtained at this level of theory. Finally, it was suggested that a qualitative series of α-lactones could be set-up by comparing the carbonyl absorption frequencies, with the proviso that such systems must be measured in comparable environments.

Table 4.3. Selected geometrical parameters for optimized structures. Bond lengths in Å, angles in degrees.

structure method	C_α–O_n	C_α–C	C–O_n	C=O	C_α–O_h	$\angle C_\alpha CO_n$(°)	ΣC_α(°)	$\angle HC_\alpha O_h H$(°)
5								
HF/6-31G(d,p)	1.479	1.438	1.305	1.171	1.340	65.0	358.2	−168.8
HF/6-31+G(d)	1.481	1.439	1.304	1.172	1.340	65.1	358.3	−169.5
MP2(fc)/6-31+G(d)	1.586	1.454	1.344	1.208	1.355	68.9	359.1	−169.1
MP2(full)/6-311G(d,p)	1.556	1.454	1.334	1.194	1.350	67.7	359.1	−167.6
AM1	1.464	1.466	1.383	1.205	1.372	61.8	357.2	+165.4
AM1/COSMO	1.487	1.463	1.378	1.212	1.365	63.0	357.9	−179.4
$\mathbf{5}_g$								
HF/6-31+G(d)	1.478	1.435	1.305	1.170	1.342	65.1		+44.7
MP2(fc)/6-31+G(d)	1.574	1.450	1.346	1.205	1.358	68.4		+35.8
AM1/COSMO	1.503	1.462	1.369	1.213	1.371	64.0		+21.2
1a								
MP2(fc)/6-31+G(d)	1.551	1.449	1.346	1.204		67.3	358.7	

Table 4.4. Calculated total energies (hartree), vibrational zero-point and thermal energies (E_{zp} and E_{th}, unscaled) for anti, **5**, and gauche, **5g**, conformers of hydroxyoxiranone along with other species involved in isodesmic equation (4.11). The energy change, ΔE_4, for eqn. 4.11 is also given at various levels of theory.

	5	**1a**	CH_4	CH_3OH	ΔE_4/kJ mol^{-1}
HF/6-31G(d,p)	−301.43366	−226.57024	−40.20171	−115.04671	−48.4
MP2(fc)/6-31+G(d)	−302.23621	−227.19009	−40.33408	−115.35784	−58.7
MP2(full)/6-311G(d,p)	−302.46599	−227.36521	−40.39804	−115.47407	−65.0
QCISD(T)(full)/6-311G(2df,p) //MP2(full)/6-311G(d,p)	−302.68557	−227.53696	−40.44666	−115.57052	−65.0
AM1	−312.39[b]				
E_{zp} /kJ mol^{-1} [HF/6-31G(d,p)]	125.59	110	124.46	144.73	
E_{th} /kJ mol^{-1} [HF/6-31G(d,p)]	11.32	9	7.50		
	5_g				
HF/6-31+G(d)	−301.43359				
MP2(fc)/6-31+G(d)	−302.23404				

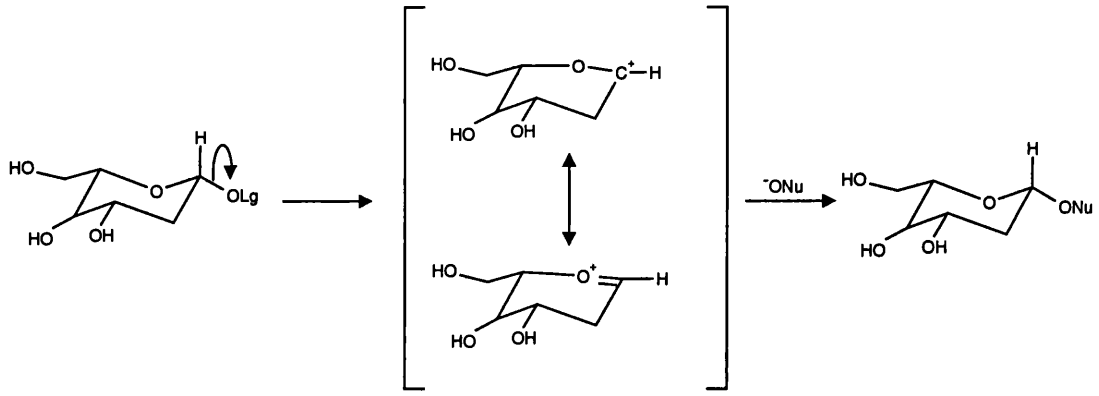
[b]AM1 energy in kJ mol^{-1}

5. INTERMEDIATES IN NON-ENZYMIC SIALOSIDE HYDROLYSIS

5.1. Introduction

A reaction intermediate is said to exist if it has a "lifetime which is long enough that the various fragments from its precursor are not still surrounded by a common solvent shell when the products are formed" (about 10^{-11}s).[68] A species with a lifetime longer than the period of a molecular vibration within the system (*cf.* around 5×10^{-13} s for a C—C stretch) has also been suggested as a criterion for the existence of an intermediate.[68,69] If such a species has too short a lifetime to become solvent equilibrated, then the mechanism is deemed to proceed through a preassociation stepwise (D_N*A_N) mechanism.[69,70]

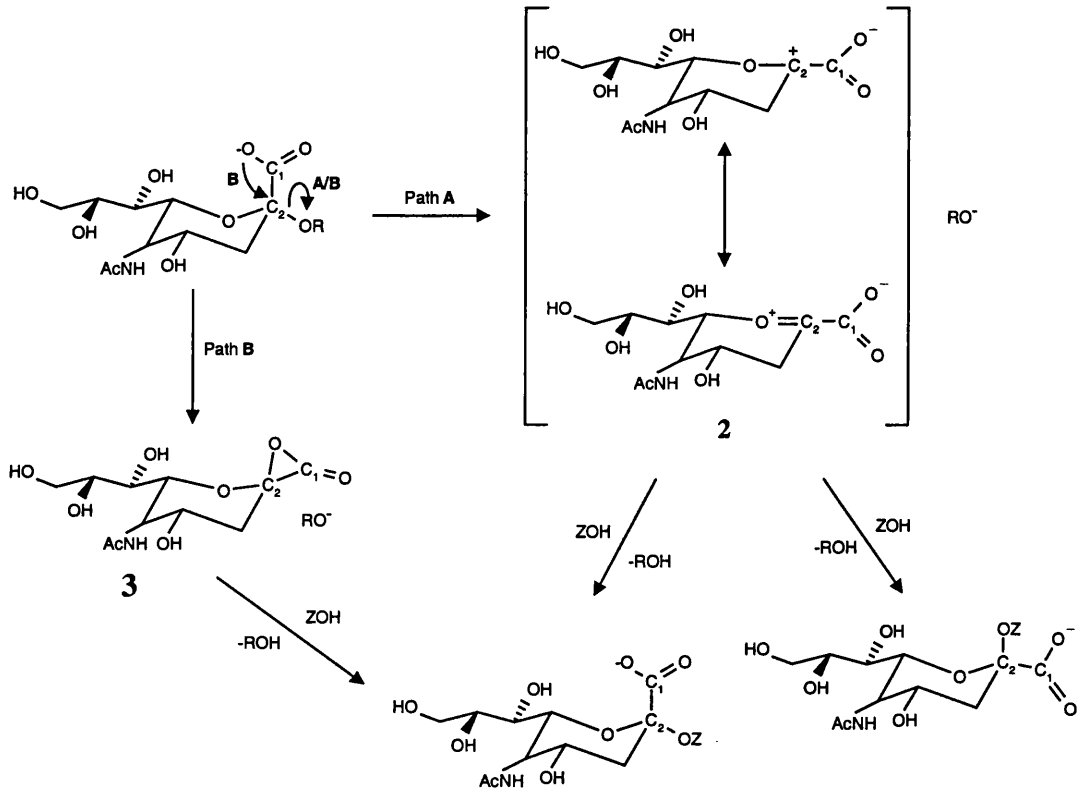
Non-enzymic catalysed glycoside hydrolysis can occur via a stepwise (D_N*A_N) nucleophilic substitution reaction that involves a high-energy oxocarbenium ion intermediate, **1** (scheme 5.1).[68,71] The lifetime of the resulting glycosyl cation has been estimated to be in the order of 10^{-12}s in an aqueous environment.[72] Conversely, in the acid catalysed hydrolysis of α-D-glucopyranosyl fluoride, the presence of anionic nucleophiles such as azide anions was found to shorten the estimated lifetime of the postulated oxocarbenium ion to about 10^{-19}s.[73] The resulting reaction products were found to be inverted and not racemised, indicative of a concerted A_ND_N mechanism.[73] However, the solvolysis of α- and β-glucopyranosyl fluorides in an ethanol/2,2,2-trifluoroethanol solvent mixture afforded racemised products for both anomers.[68] This indicated that solvolysis of the glycoside may have proceeded via a mechanism that involved an oxocarbenium ion intermediate. The different retained/inverted product ratios for a series of leaving groups suggested the oxocarbenium ion was not solvent equilibrated, but existed in the presence of the nucleophile as a solvent separated ion pair.[68]



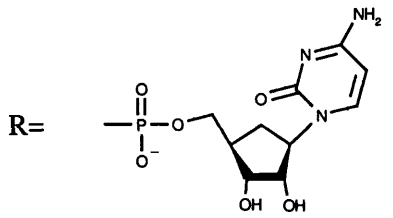
1

Scheme 5.1

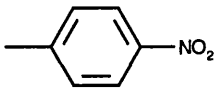
Sialyl oxocarbenium ions (sialyl zwitterions), **2** are more stable than the parent glycosyl cations. Their lifetime in aqueous solvent has been estimated to be around 3 x 10^{-11}s, which is 30 times greater than the estimated lifetime of the glycosyl cation.[74] This is long enough to fully satisfy both stipulated criteria for the existence of an intermediate. Further, the solvolysis of sialosides with a CMP leaving group, **4** gave a "nearly equimolar" product solution of α- and β-methyl glycosides of sialic acid (Scheme 5.2, Path A, Z=Me).[75] The racemic product mixture points evidence towards a solvent equilibrated sialyl zwitterion.



Scheme 5.2

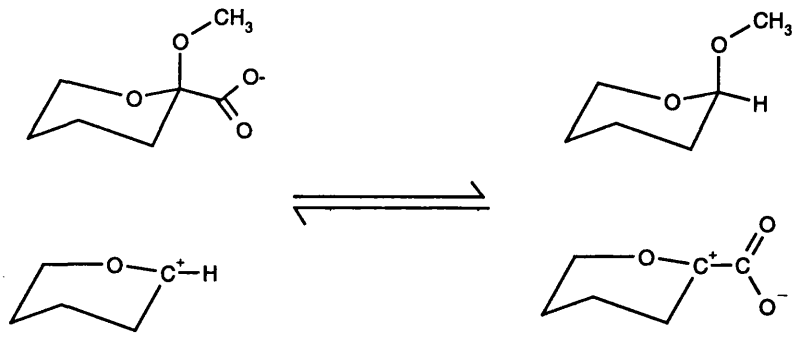


4



5

The extra stabilisation of the sialyl zwitterion, **2**, over the parent glycosyl cation has been attributed to the through-space electrostatic stabilisation of the positively charged centre by the adjacent carboxylate group.[76] Indeed, an isodesmic relationship involving both species (Scheme 5.3) was exothermic by 460 kJ mol^{-1} in the gas phase, and 71 kJ mol^{-1} when electrostatic continuum solvation effects were taken into account.[76]

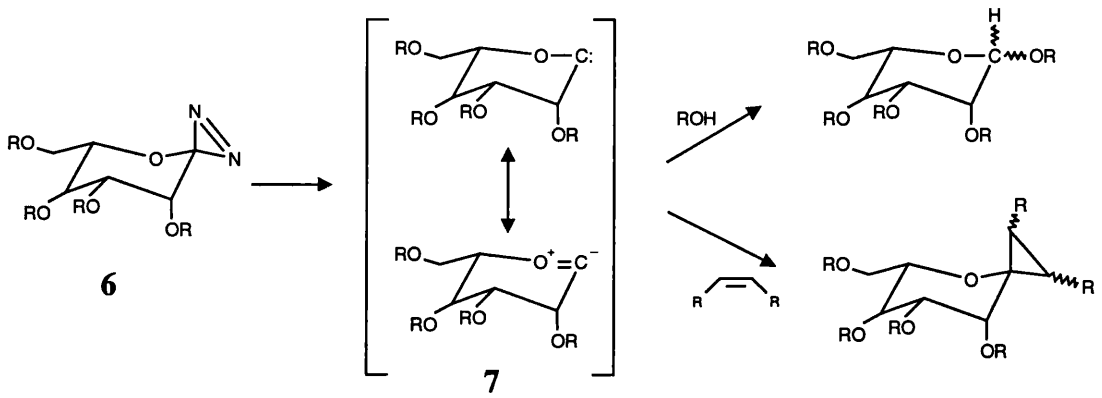


Scheme 5.3

Although electrostatic effects are important in stabilising a positively charged centre in the sialyl zwitterion, covalent interactions with the carboxylate group could also stabilise the reaction centre through NGP, leading to the formation of an α-lactone intermediate, **3**. The non-enzymic hydrolysis of *p*-nitrophenyl-N-acetyl-α-D-neuraminic acid, **5**, has been suggested to take place via an α-lactone intermediate, aiding the departure of the *p*-nitrophenyl group (scheme 5.2, path B, Z=H).[1] This mechanism has been suggested on the basis of an inverse solvent isotope effect as well as a larger β-deuterium KIE found at lower pH conditions, under which the carboxylate group is less likely to covalently participate.[1] NGP by the carboxylate group was discounted for the solvolysis of CMP-Neu5ac because such a mechanism would only lead to a retained product.[75]

The secondary ^{14}C KIEs at the carboxylate carbon (C1) have also been measured[77] and calculated.[76] The measured KIE was found to have a normal value of 1.013, indicating a degree of loosening of the environment around the C1 carbon.[77] Furthermore, theoretical studies have shown that on the departure of the leaving group, the carboxylate O—C—O angle increases along with an accompanying elongation of the C1—C2 bond.[76] Such geometrical changes would be consistent with a normal isotope effect. Much larger changes in these geometrical parameters would represent the decarboxylation of a sialyl zwitterion to yield a sialyl glycosylidene carbene and carbon dioxide.

Interestingly, geometry optimisations at HF/3-21G of protonated pyruvate, $CH_3C^+(OH)COO^-$, in the gas phase resulted in the spontaneous decarboxylation of the molecule.[78] Initial studies by the author of this thesis involved optimisation of the sialyl zwitterion in *vacuo* using the AM1 level of theory within the SPARTAN program.[90] This intermediate also underwent spontaneous decarboxylation to yield an associated complex of the sialyl glycosylidene carbene and CO_2. Glycosylidene carbenes, **7** have been generated experimentally from diazirines, **6** and have been used as glycosylating agents.[79] The lifetime of such carbenes has been estimated to be around 10^{-6}s in acetonitrile solvent.[80] Glycosylidene carbenes undergo standard reactions of carbenes, such as O—H bond insertion[79] and addition to unsaturated double bonds[81] (scheme 5.4).



Scheme 5.4

In this chapter, the relative stability of the three possible intermediates discussed above are investigated using two separate models. **Model I** incorporates the pyranose oxygen, anomeric carbon (C2) and carboxylate group to define the three intermediates (figure 5.1A). **Model II** includes the full pyranose ring with the carboxylate group (figure 5.1B). The stabilising effects which exist in addition to the through–space electrostatic stabilisation by the carboxylate group are investigated, along with their effects on the relative stability of the three intermediates.

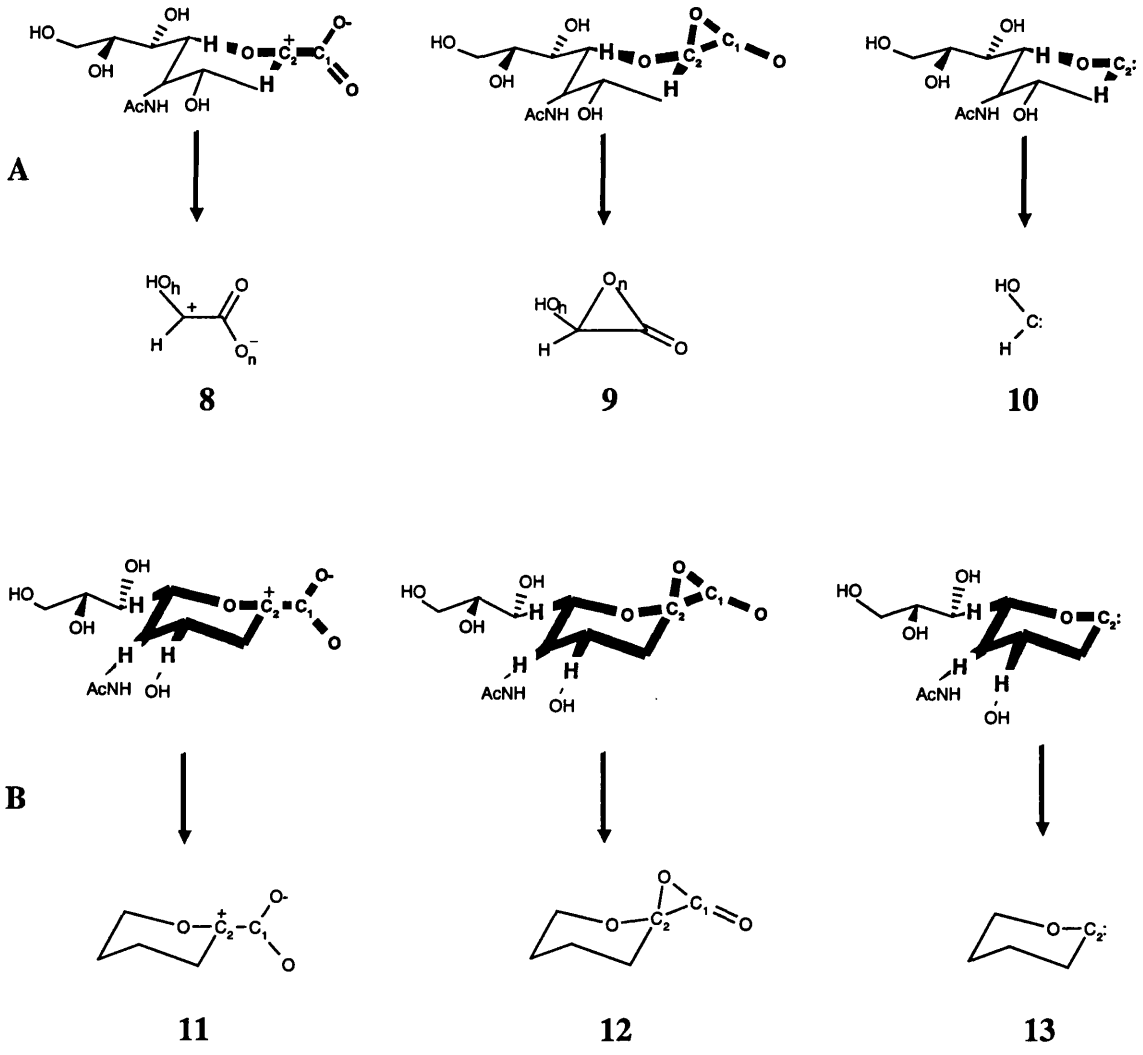


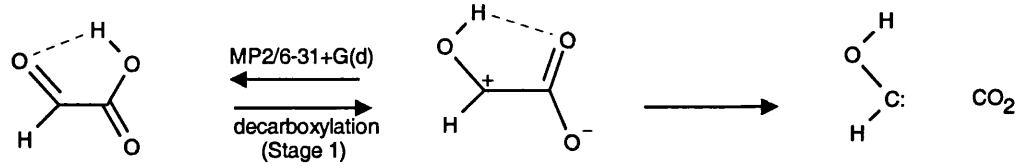
Fig 5.1. The relationship of model I (A) and model II (B) intermediates to the full sialoside system.

After discussing the structures for model 1, the stabilising effect of the hydroxyl substituent on the electron deficient centre is then examined, followed by the effect of anchoring model I into a pyranose ring. The relative stability of the three

intermediates in the gas phase are then investigated, and the effect of solvation upon the relative stability of the zwitterion, α-lactone and carbene is then discussed. The chapter is then summarised with a conclusion.

5.2. Structure and Stability of Model I Intermediates

Hydroxy-zwitterion (8). Unconstrained optimisation of **8** at HF/6-31+G(d) yields a planar structure with an intramolecular hydrogen bond between the hydroxyl hydrogen and a carboxylate oxygen, where the C_α—O_hH bond is anti to C_α—H. However, this structure is found to undergo a spontaneous intramolecular proton transfer to give glyoxylic acid (Scheme 5.5, **14**) on the MP2(fc)/6-31+G(d) energy surface. The reverse process has been investigated by Bock and Redington[82] as the first step in the decarboxylation of **14** (Scheme 5.5), but their theoretical study used structures only obtained at the Hartree-Fock level.



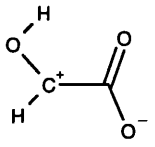
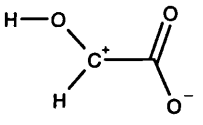
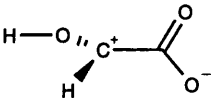
14

Scheme 5.5

The conformation of **8** in which the C_α—O_hH bond is syn to the C_α—H bond is a more appropriate analogue for the sialyl zwitterion, **2**. This conformation is found to be a first-order saddle point on the HF/6-31+G(d) and MP2/6-31+G(d) potential energy surfaces (Table 5.6). It represents a transition structure for rotation about the C_α–O_h bond, but is stable with respect to in-plane distortions. Interestingly, Bock and Redington[82] described the anti conformer of **8** as a weakly associated complex of hydroxycarbene and carbon dioxide with a C_α–C distance of 3.0 Å at the HF/6-31G(d) level. The syn conformer of **8** has a perfectly normal C_α–C bond, the length of which is reduced by electron correlation from 1.598 Å [HF/6-31+G(d)] to 1.523 Å [MP2(fc)/6-31+G(d)].

As well as having the carboxylate group in plane with the molecular system as in $\mathbf{8}_{\text{syn-planar}}$, the group can also adopt a conformation where it is co-planar with the electrophilic centre, $\mathbf{8}_{\text{syn-twist}}$. This was also found to be a first-order saddle point on the HF/6-31+G(d) energy surface for the formation of the corresponding α-lactone. However, $\mathbf{8}_{\text{syn-twist}}$ could not be located on the MP2/6-21+G(d) surface; attempts to find such a species resulted in the collapse of the zwitterion to yield α-lactone, $\mathbf{9}_{\text{gauche}}$. The relative energies of the three zwitterion conformations at HF/6-31+G(d) are given in table 5.1.

Table 5.1. Relative energies (kJ mol^{-1}) of the three reported hydroxy-zwitterion conformers at HF/6-31+G(d).

Conformer	$\mathbf{8}_{\text{anti}}$	$\mathbf{8}_{\text{syn-planar}}$	$\mathbf{8}_{\text{syn-twist}}$
Relative Energy /kJ mol^{-1}	-73	0	-14

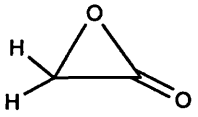
The twist hydroxy-zwitterion is an appropriate model for the sialyl zwitterion in which the carboxylate group is co-planar with the pyranose ring. It has been suggested that both planar and twist conformations exist in solution under different circumstances.[74] When the sialyl zwitterion is solvent equilibrated, the lower energy twist conformation may be adopted. However, in the presence of anions, the zwitterion may assume the planar conformation to avoid any electrostatic repulsion between the carboxylate group and an adjacent nucleophile.[74] It is also noted that this planar conformation is also adopted when the sialic acid moiety is bound to the active site of the sialidase enzyme (chapter 7).

*Hydroxyoxiranone (**9**).* The rotamers of hydroxyoxiranone have already been discussed (section 4.2). Briefly, the gauche conformer lies 5.7 kJ mol^{-1} above the anti conformer on the MP2/6-31+G(d) energy surface and is the more appropriate model for the sialyl lactone, **3**. These were the only rotamers obtained at the HF/6-31+G(d) and MP2/6-31+G(d) levels of theory.

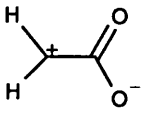
Hydroxy-carbene (10). Two conformers of the singlet hydroxy-carbene have been located on the MP2/6-31+G(d) energy surface. The syn conformer is the most appropriate model for a sialyl glycosylidene carbene and is located 29 kJ mol^{-1} above the anti form, in reasonable agreement with other groups.[83,84] Pople *et al.* have calculated a high rotational barrier (ca. 120 kJ mol^{-1}) associated with the inter-conversion of the syn and anti conformers.[85] The high barrier is consistent with the rotation around a partial double bond (table 5.6).

5.3. The Stabilising Effect of the Pyranose Oxygen

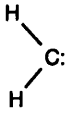
A possible stabilising effect upon the electron deficient centre in all three intermediates is the n_π-p_π interaction that involves a lone pair of electrons on the adjacent pyranose oxygen. The magnitude of this stabilising effect can be estimated by comparing the model I intermediates, with the relative unsubstituted analogues **15-17**.



15



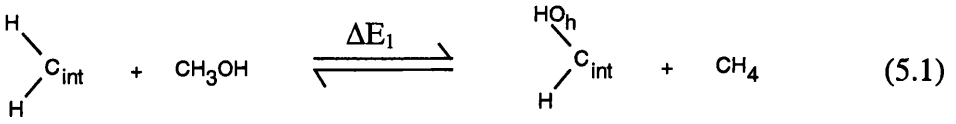
16



17

For the purposes of the following discussion, these unsubstituted species will be generically referred to as $HC_{int}H$. Likewise, model I intermediates will be referred to as $HOC_{int}H$. As noted in section 5.2, The gauche conformer of hydroxyoxiranone is used, along with the syn-planar zwitterion. The syn-twist zwitterion is not compared with the unsubstituted model because the corresponding form of **16** could not be located on either the HF/6-31+G(d) or the MP2/6-31+G(d) energy surface. Any attempt to do so resulted in its collapse to form oxiranone, **15**. This is consistent with the results of Antolovic *et al.* in which a similar optimisation at HF/3-21+G(d) also yielded **15**.[86] However, the planar conformation of **16** is found to be a first-order saddle point on both the HF/6-31+G(d) and MP2/6-31+G(d) energy surface which corresponds to a TS structure for the rotation about the C_α—C bond.

$HOC_{int}H$ can be related to $HC_{int}H$ through the following isodesmic equation:



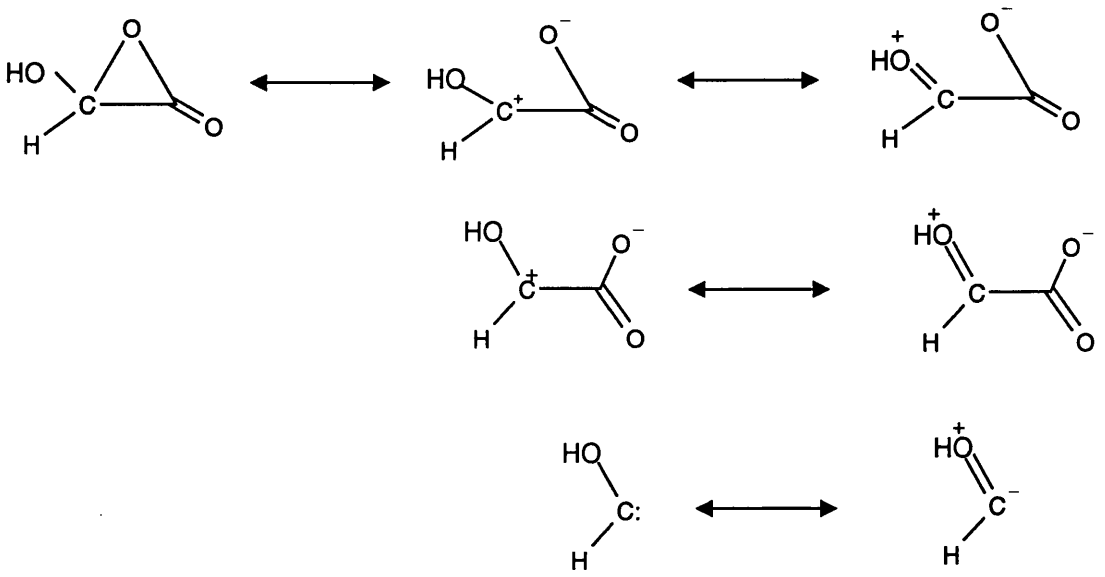
(5.1)

The calculated value of ΔE_1 for each intermediate is summarised in table 5.2. The total energies for all intermediates are given in tables 5.8-5.10.

Table 5.2. The energy change, ΔE_1 (eqn. 5.1), calculated at MP2(fc)/6-31+G(d), along with the C_α—O_h bond length for each intermediate.

	α-Lactone	Zwitterion	Carbene
ΔE_1/kJ mol^{-1}	-53.0	-204.7	-190.8
C_α—O_h/Å	1.358	1.280	1.320

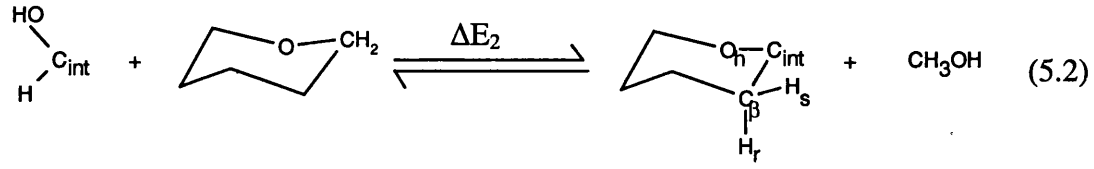
The introduction of the hydroxyl substituent provides a degree of conjugative stabilisation in all three intermediates, as noted by the exothermic energy change in eqn. 5.1. The gauche α-lactone is less strained as compared with the unsubstituted species, while the electron deficient carbene is stabilised to a greater extent, as evidenced by an increased n_π-p_π interaction (i.e. a reduction in the C_α—O_h bond). However, it is the positively charged centre of the zwitterion that benefits most from the donation of electrons into the empty p-orbital, giving the most exothermic energy change of 205 kJ mol^{-1}. The conjugative stabilisation of all three intermediates can also be represented as resonance structures (scheme 5.6).



Scheme 5.6

5.4. Effect of Ring Restraints on Model I Intermediates

The energy change associated with the following isodesmic relationship (eqn. 5.2) is given in table 5.3, along with the C—O_h, and C_β—H_r bond lengths found in all three model II intermediates.



(5.2)

Model I **Model II**

Table 5.3. The energy change associated with equation 5.2 at MP2/6-31+G(d)//HF/6-31+G(d), along with the geometrical C_α—O_h and C_β—H_r parameters for each model II intermediate.

	α-Lactone	Zwitterion	Carbene
ΔE_2/ kJ mol^{-1}	-24.9	-62.8	-26.5
C_α—O_h/Å	1.344	1.254	1.294
C_β—H_r/Å	1.087	1.089	1.092

The energy change for eqn. 5.2 gives the energy associated with restraining model I into a pyranose ring (Model II). For all intermediates, this change is exothermic. The observed stabilising effect may originate from the combination of three possible phenomena:

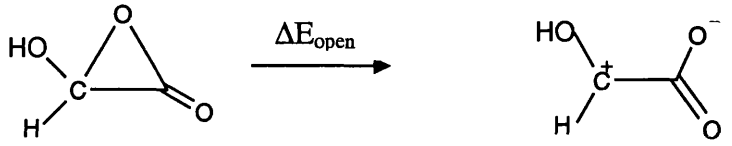
1). A slight conformation change in Model I compared to Model II.
2). A change in the magnitude of the first-order conjugation taking place.
3). The introduction of hyperconjugation at the β-centre.

The α-lactone stability is affected by the ring restraint: the corresponding dihedral angle to $HC_{\alpha}O_{h}H$ in $\mathbf{9}_{\mathbf{gauche}}$ is some 9° larger in the spiro-fused α-lactone. This slight conformation change of the α-lactone in the pyranose ring system over the smaller model allows some reduction in the ring strain to occur. This is evidenced by an elongation of the C_{α}—O_{n} bond in **12** (by 0.032Å). Adam's work on the spiro-fused α-lactones gave an indication that the strain associated with these systems is inversely proportional to the size of the adjacent ring.[35] In addition to the conformational change between the smaller and larger models, the C—O_{h} bond is 0.002Å longer, suggestive of a slight weakening of the conjugative stabilisation by the oxygen atom in the model II α-lactone.

While there seems to be some evidence for the reduction in the C_{α}—O_{h} interaction in the model II α–lactone, the converse is true for the zwitterion and carbene intermediates. The C_{α}—O_{h} distance in both species is shortened by 0.004Å as compared to the distance in Model I. There is also some indication of hyperconjugation in these two systems: the C_{β}—H_{r} bond length is 0.002Å and 0.005Å longer for the zwitterion and carbene respectively, as compared with the unsubstitued pyran molecule (where C_{β}—H_{r} is 1.087Å). In addition, it is also noted that for both intermediates, there is a smaller conformation change (<3°) between Model I and Model II as compared with the corresponding change in the α-lactone intermediate.

5.5. Relative Energies of α-Lactone, Zwitterion and Carbene

The relative energies of the three intermediates are given in table 5.4. *In vacuo*, the energy change, ΔE_{open}, for the process:



(5.3)

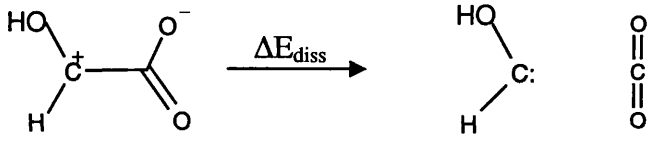
is positive. This is the case at all levels of theory, but the effect of electron correlation is to increase the value of ΔE_{open}, by some 30 kJ mol^{-1}. Thus the α-lactone, $\mathbf{9}_{gauche}$ is found to be more stable than the 1,3-dipolar system despite its highly strained structure. At the AM1 semi-empirical level of theory, the only α-lactone conformer found was $\mathbf{9}_{anti}$, which was 60.4 kJ mol^{-1} lower in energy than the syn-planar zwitterion. The energy change for the ring opening of oxiranone, **15**, to the unsubstituted zwitterion is calculated to be 262 kJ mol^{-1} at MP2(fc)/6-31+G(d), which is more endothermic than that obtained for the substituted model at the same level of theory. The syn-twist hydroxy-zwitterion obtained at the HF/6-31+G(d) level is found to be slightly more stable with respect to $\mathbf{9}_{gauche}$ compared to its syn-planar conformer by some 14 kJ mol^{-1} and 27 kJ mol^{-1} using the HF/6-31+G(d) and MP2/6-31+G(d)//HF/6-31+G(d) levels of theory respectively.

Using Model II, ΔE_{open} is calculated as 21 kJ mol^{-1} at HF/6-31+G(d). Accounting for electron correlation increases ΔE_{open} to 46 kJ mol^{-1}. This result is in good agreement with those of Horenstein, who calculated the same relative energy to be 21 kJ mol^{-1} and 47 kJ mol^{-1} at HF/6-31G(d) and MP2/6-31G(d)//HF/6-31G(d) respectively.[74] However, in stark contrast, at the AM1 level of theory, the model II zwitterion is found to be 10.3 kJ mol^{-1} lower in energy than the α-lactone.

Table 5.4. Energies associated with α-lactone ring opening (ΔE_{open}) and zwitterion dissociation (ΔE_{diss}) *in vacuo*. The former process results in the zwitterion intermediate while the latter generates a singlet carbene and CO_2. BSSE corrected ΔE_{diss} also given.

Method	ΔE_{open}/kJ mol^{-1}	ΔE_{diss}/kJ mol^{-1}	BSSE corrected ΔE_{diss}/kJ mol^{-1}
Model I-Unsubstituted Intermediates			
MP2(fc)/6-31+G(d)	261.9	-8.8	
Model I-Comparison with syn-planar zwitterion			
HF/6-31+G(d)	81.2	–44.2	–52.8
HF/6-311+G(3df,2p) //HF/6-31+G(d)		–56.8	
MP2(fc)/6-31+G(d) //HF/6-31+G(d)	110.7	3.4	
MP2(fc)/6-31+G(d)	110.2	5.1	–25.7
AM1	60.4	-16.4	
Model I-Comparison with syn-twist zwitterion			
HF/6-31+G(d)	67.6	-30.5	-39.1
MP2(fc)/6-31+G(d) //HF/6-31+G(d)	83.6	30.9	-0.3
Model II-Comparison with twist zwitterion			
HF/6-31+G(d)	20.7	15.5	8.7
MP2(fc)/6-31+G(d) //HF/6-31+G(d)	45.5	66.9	36.8
AM1	-10.3	-10.0	

In vacuo, the energy change for model I associated with the process:



(5.4)

was found to be exothermic at both the semi-empirical AM1 level and the HF/6-31+G(d) level of theory. The inclusion of electron correlation seems to preferentially stabilise the zwitterion compared to the infinitely separated dissociated products. However, the method used to describe the process given in eqn. 5.4 results in the preferential stabilisation of the zwitterion over the fragments. This is due to the larger number of basis functions available in the super-molecule (the zwitterion) to describe each individual fragment within it. This effect originates from the use of incomplete basis sets and is termed *Basis Set Superposition Error (BSSE)* and has no physical significance.

A modified[87] Boys-Bernardi counterpoise approach[88] can be used to estimate the amount of stabilisation afforded to the zwitterion through BSSE. Firstly, the energy of each fragment is calculated, in which both species assume their super-molecule geometry. The resulting sum is designated $\varepsilon_{frag.}$ Secondly, the energy of each fragment in the super-molecule, E_{frag}, is calculated in the presence of the counter fragment's molecular orbitals (often termed ghost orbitals). Thus:

$$BSSE = E_{frag}\text{-}\varepsilon_{frag} \qquad (5.5)$$

The value associated with BSSE is then subtracted from the energy of the super-molecule (in this case 8.6 kJ mol^{-1} and 30.8 kJ mol^{-1} at HF/6-31+G(d) and MP2/6-31+G(d) levels of theory respectively for the model I syn-planar zwitterion). This gives a corrected energy change for the process in eqn 5.4. While BSSE has a small effect on the ΔE_{dis} at Hartree-Fock level, the corrected MP2/6-31+G(d) result swings the syn-planar zwitterion dissociation from being an endothermic (5.1 kJ mol^{-1}) to being an exothermic process (-25.7 kJ mol^{-1}).

BSSE can also be accounted for by using a larger, more flexible (and more complete) basis set. Single-point energies at HF/6-311+G(3df,2p) on the HF/6-31+G(d) structure yield a dissociation energy of -56.8 kJ mol^{-1}, compared to the BSSE corrected -52.8 kJ mol^{-1}. Therefore, even after accounting for basis set effects, electron correlation preferentially stabilises the zwitterion with respect to the

dissociation products, although decarboxylation is still a thermodynamically favourable process.

Therefore, for the hydroxyl substituted model in which the zwitterion adopts the syn-planar conformation, the decarboxylation process is found to be exothermic at all levels of theory after accounting for BSSE. For the unsubstituted model, zwitterion, **16**, is higher in energy than the dissociated fragments by 8.8 kJ mol^{-1}, before allowing for the BSSE correction.

Considering Model I in which the zwitterion now adopts the syn-twist conformation, there is an exothermic energy change associated with the formation of hydroxycarbene and carbon dioxide through the decarboxylation of $\mathbf{8}_{\text{syn-twist}}$. However, at MP2(fc)/6-31+G(d), the energy change is only -0.3 kJ mol^{-1} as opposed to –25.7 kJ mol^{-1} for the decarboxylation of the syn-planar species. Thermodynamically, the dissociation of the syn-planar zwitterion may be more important than for its syn-twist analogue (whereas the syn-twist conformer would favour the exothermic ring closure process to form the α-lactone).

In model II, the greater stabilising effect of the pyranose ring on the zwitterion species over the carbene leads to an endothermic process for the decarboxylation of the zwitterion, even after applying the BSSE correction. However, using the AM1 Hamiltonian, decarboxylation to infinitely separated products is predicted to be an exothermic process, in agreement will the smaller model I, but in disagreement with the *ab initio* results obtained for model II.

5.6. Relative Energies of the Three Intermediates in Solution

Table 5.5 gives the relative energy changes in solution between the zwitterion and the α-lactone (ΔE_{open}) and carbene (ΔE_{diss}) intermediates. In the presence of solvent, localised charges in the 1,3-dipolar system should be greatly stabilised, compared to the highly strained α-lactone. It is therefore surprising to note that the model I α-lactone is lower in energy than both zwitterion conformers (ring opening to form the syn-twist zwitterion is 5.3 kJ mol^{-1} less endothermic). Conversely, the larger model II predicts the zwitterion to be the more stable intermediate in solution, with the ring

opening process being exothermic by 8.2 kJ mol^{-1} at MP2/6-31+G(d)//HF/6-31+G(d). These results suggest that under the restriction of using *in vacuo* geometries with the IPCM method, the smaller model under-estimates the relative stability of the zwitterion as compared with the α-lactone in solution. Interestingly, using the AM1/COSMO method to describe both model I and model II, the zwitterion is found to be lower in energy. At this level of theory, geometry optimisation is available within the continuum dielectric. Horenstein reported geometry optimisations within a dielectric continuum using the self-consistent field isodensity polarised continuum model (SCIPCM) at HF/6-31G(d). At this level of theory, the zwitterion was 42.3 kJ mol^{-1} lower in energy than the spiro-fused α-lactone. Single-point energy calculations on the SCIPCM structures at B3LYP/6-31G(d) and MP2/6-31G(d), also reported the zwitterion to be lower in energy, with values of ΔE_{open} of -32.6 kJ mol^{-1} and -7.5 kJ mol^{-1} respectively.[76] This would suggest the AM1/COSMO value of around 120 kJ mol^{-1} obtained by Gillot[89] over-estimates the energy associated with the ring opening of the α-lactone compared with *ab initio* levels of theory.

The dissociation of the zwitterion to form the carbene and carbon dioxide is endothermic at all levels of theory. Although the *ab initio* results have not been corrected for BSSE (ca 30 kJ mol^{-1} when using MP2 methods), this effect is unlikely to alter the overall qualitative result. Electrostatic solvent effects therefore render the decarboxylation process unfavourable.

Table 5.5. Energies associated with α-lactone ring opening (ΔE_{open}) and zwitterion dissociation (ΔE_{diss}) in solution. The former process results in the zwitterion intermediate while the latter generates a singlet carbene and CO_2. BSSE corrected ΔE_{diss} also given.

Method	ΔE_{open}/kJ mol^{-1}	ΔE_{diss}/kJ mol^{-1}
Model I-Comparison with syn-planar zwitterion		
MP2(fc)/6-31+G(d) //MP2/6-31+G(d)	21.5	84.5
MP2(fc)/6-31+G(d) //HF/6-31+G(d)	17.5	82.7
AM1/COSMO	-101.0	111.8
Model I-Comparison with syn-twist zwitterion		
MP2(fc)/6-31+G(d) //HF/6-31+G(d)	12.2	88.0
AM1/COSMO	-103.7	114.5
Model II-Comparison with twist zwitterion		
MP2(fc)/6-31+G(d) //HF/6-31+G(d)	-8.2	107.7
AM1/COSMO	-119.4	66.6

5.7. Conclusions

Although, $\mathbf{8}_{\text{syn-twist}}$, $\mathbf{8}_{\text{syn-planar}}$, $\mathbf{9}_{\text{gauche}}$ and $\mathbf{10}_{\text{syn}}$ are not the lowest energy conformers of their respective systems, they are the most appropriate models for the description of the sialoside intermediates of interest. Introduction of an adjacent oxygen atom as compared with unsubstituted analogues affords a degree of stabilisation through n_{π}-p_{π} bonding interactions. This is most effective for the positively charged centre of the zwitterion (-205 kJ mol^{-1}) and the electron-deficient carbene centre (-191 kJ mol^{-1}). Further stabilisation is obtained by restraining the system into a pyranose ring. This may originate from either additional first- and/or second-order conjugative effects and/or slight conformational changes in the pyranose system as compared to

the conformations used in the smaller models. At *ab initio* levels of theory, the model II zwitterion is less stable than the α-lactone *in vacuo*, but more stable in solution, while decarboxylation to form the carbene and CO_2 is unfavourable *in vacuo* and in solution. However, using the AM1 hamiltonian for model II, the zwitterion is predicted to be more stable than the α-lactone by some 10 kJ mol^{-1} but less stable than the infinitely separated carbene and CO_2. In solution, the semi-empirical result agrees qualitatively with those obtained at *ab initio* levels of theory. However, there is evidence to suggest that the AM1/COSMO method may be over estimating the stabilisation of the zwitterion in solution as compared to published *ab initio* results. The results in solution given in this chapter only provide a degree of stabilisation from a medium with a given dielectric and do not give any information on the stabilising effects arising through specific intermolecular interactions between solute and solvent.

Table 5.6. Selected geometrical parameters for the α-lactone, zwitterion and carbene intermediates

Structure Method	C_α–O_n	C_α–C	C–O_n	C=O	C_α–O_h	$\angle C_\alpha CO_n$(°)	ΣC_α(°)	$\angle HC_\alpha O_h H$(°)	$\nu^\ddagger$/cm^{-1}
9_{anti}									
AM1	1.464	1.466	1.383	1.205	1.372	61.8	357.2	+165.4	
9_{gauche}									
HF/6-31+G(d)	1.478	1.435	1.305	1.170	1.342	65.1		+44.7	
MP2(fc)/6-31+G(d)	1.574	1.450	1.346	1.205	1.358	68.4		+35.8	
AM1/COSMO	1.503	1.462	1.369	1.213	1.371	64.0		+21.2	
15									
MP2(fc)/6-31+G(d)	1.551	1.449	1.346	1.204		67.3	358.7		
12									
HF/6-31+G(d)	1.503	1.437	1.303	1.173	1.344	66.3	358.1	+53.8	
AM1	1.493	1.467	1.364	1.205	1.386	63.5	357.6	+48.3	
AM1/COSMO	1.512	1.466	1.362	1.212	1.381	64.5	357.8	+42.2	
$8_{syn\text{-}planar}$									
HF/6-31+G(d)		1.598	1.213	1.203	1.250	107.4	360.0	0.0	118*i*
MP2(fc)/6-31+G(d)		1.523	1.267	1.247	1.280	107.1	360.0	0.0	183*i*
AM1		1.651	1.235	1.229	1.286	108.1	360.0	0.0	82*i*
AM1/COSMO		1.522	1.264	1.262	1.286	117.1	360.0	0.0	30*i*

$8_{syn\text{-}twist}$								
HF/6-31+G(d)	1.543	1.210	1.210	1.258	110.2	360.0	0.0	240*i*
AM1	1.634	1.232	1.232	1.289	110.4	360.0	0.0	100*i*
AM1/COSMO	1.519	1.263	1.263	1.286	118.5	360.0	0.0	
16								
HF/6-31+G(d)	1.586	1.211	1.211		110.6	360.0		453*i*
MP2(fc)/6-31+G(d)	1.445	1.281	1.281		113.1	360.0		722*i*
11								
HF/6-31+G(d)	1.542	1.218	1.213	1.254	109.4	360.0	2.9	
AM1	1.640	1.234	1.234	1.294	111.1	360.0	1.4	
AM1/COSMO	1.532	1.259	1.259	1.296	117.9	360.0	1.6	
10_{syn}								
HF/6-31+G(d)				1.298			0.0	
MP2(fc)/6-31+G(d)				1.320			0.0	
AM1				1.280			0.0	
AM1/COSMO				1.282			0.0	
13								
HF/6-31+G(d)				1.294			2.5	
AM1				1.281			1.2	
AM1/COSMO				1.286			1.3	

Table 5.7. Total energies (Hartrees) for methane, methanol and pyran molecules used to complete isodesmic relationships 5.1 and 5.2.

Method	CH_4	CH_3OH	$C_5H_{10}O$
MP2/6-31+G(d) //HF/6-31+G(d)		-115.35651	-270.85708
MP2/6-31+G(d)	–40.33408	–115.35784	

Table 5.8. Total energies for obtained α-lactone structures. Results calculated using *ab initio* levels of theory are given in Hartrees while AM1 and AM1/COSMO results are reported in kJ mol^{-1}.

Method	$\mathbf{9}_{anti}$	$\mathbf{9}_{gauche}$	**15**	**12**
HF/6-31+G(d)		–301.43359		-456.42234
MP2(fc)/6-31+G(d) //HF/6-31+G(d)		-302.22633		-457.73637
MP2(fc)/6-31+G(d)		–302.23404	–227.19009	
IPCM-MP2(fc)/6-31+G(d) //HF/6-31+G(d)		–302.24251		-457.74792
IPCM-MP2(fc)/6-31+G(d)		–302.25167		
AM1	–312.39			-334.19
AM1/COSMO		–373.83		-385.33

Table 5.9. Total energies for obtained zwitterion structures. Results calculated using *ab initio* levels of theory are given in Hartrees while AM1 and AM1/COSMO results are reported in kJ mol^{-1}.

Method	$\mathbf{8}_{\text{syn-planar}}$	$\mathbf{8}_{\text{syn-twist}}$	**16**	**11**
HF/6-31+G(d)	–301.40265	-301.40784	–226.48751	-456.41448
HF/6-311+G(3df,2p) //HF/6-31+G(d)	–301.51497			
MP2(fc)/6-31+G(d) //HF/6-31+G(d)	–302.18423	-302.19453		-457.71903
MP2(fc)/6-31+G(d)	–302.19206		–227.09034	
IPCM-MP2(fc)/6-31+G(d) //HF/6-31+G(d)	–302.23585	-302.23788		-457.75102
IPCM-MP2(fc)/6-31+G(d)	–302.24350			
AM1	–251.96	-260.72		-344.47
AM1/COSMO	–474.82	-477.49		-504.68

Table 5.10. Total energies for obtained carbene and CO_2 structures. Results calculated using *ab initio* levels of theory are given in Hartrees while AM1 and AM1/COSMO results are reported in kJ mol^{-1}.

Method	$\mathbf{10}_{\text{syn}}$	**17**	**13**	CO_2
HF/6-31+G(d)	–113.78069		-268.76977	–187.63879
HF/6-311+G(3df,2p) //HF/6-31+G(d)	–113.82739			–187.70923
MP2(fc)/6-31+G(d) //HF/6-31+G(d)	–114.07073		-269.58137	–188.11219
MP2(fc)/6-31+G(d)	–114.07216	–38.97572		–188.11796
IPCM-MP2(fc)/6-31+G(d) //HF/6-31+G(d)	-114.08824		–269.59387	–188.11611
IPCM-MP2(fc)/6-31+G(d)	–114.08934			–188.12130
AM1	65.83		-20.23	–334.14
AM1/COSMO	8.99		-66.12	–372.00

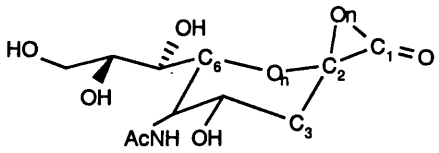
Table 5.11. Counterpoise correction data for both Model I and Model II fragments (in Hartrees).

Method	$\mathbf{10}_{syn}$	**17**	**13**	CO_2
fragments distorted as in $\mathbf{8}_{syn\text{-}planar}$:				
HF/6-31+G(d)	−113.77328			−187.57060
MP2(fc)/6-31+G(d)	−114.06505			−188.05749
fragments distorted as in $\mathbf{8}_{syn\text{-}planar}$ with ghost orbitals of counter-fragment:				
HF/6-31+G(d)	−113.77575			−187.57142
MP2(fc)/6-31+G(d)	−114.07326			−188.06099
fragments distorted as in $\mathbf{8}_{syn\text{-}twist}$:				
HF/6-31+G(d)	−113.77424			−187.57053
MP2(fc)/6-31+G(d) //HF/6-31+G(d)	−114.06259			−188.07256
fragments distorted as in $\mathbf{8}_{syn\text{-}twist}$ with ghost orbitals of counter-fragment:				
HF/6-31+G(d)	−113.77678			−187.57077
MP2(fc)/6-31+G(d) //HF/6-31+G(d)	−114.07107			−188.07595
fragments distorted as in **13**:				
HF/6-31+G(d)			-268.76362	−187.56272
MP2(fc)/6-31+G(d) //HF/6-31+G(d)			-269.57474	-188.06758
fragments distorted as in **13** with ghost orbitals of counter-fragment:				
HF/6-31+G(d)			-268.76498	−187.56387
MP2(fc)/6-31+G(d) //HF/6-31+G(d)			-269.58172	-188.07206

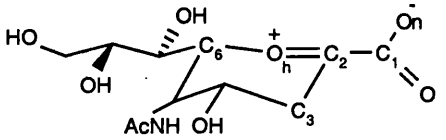
6. EXPLICIT SOLVATION OF SIALYL α-LACTONE AND ZWITTERION INTERMEDIATES

6.1. Introduction

Solvation effects upon the relative stability of possible intermediates (including the sialyl α-lactone, **1** and sialyl zwitterion, **2**) formed in the non-enzymic catalysed hydrolysis of sialosides were introduced in chapter 5. The simplified tetrahydryopyran model incorporated the pyranose ring and the carboxylate group, but excluded all other side chains in the sialoside system. Solvation effects were modelled using continuum solvation techniques with a constant dielectric (78.4 for water). While such methods are computationally efficient and give an indication of the solvent electrostatic stabilisation upon the solute, any specific solvent-solute interactions, by inference, are omitted.



1



2

To account for specific solute-solvent interactions requires the incorporation of explicit water molecules into the simulation. While QM methods enable simulation of chemical reactivity, they are computationally expensive. Thus describing a solute surrounded by a large number of water molecules (*cf.* 500) would be intractable if modelled at *ab initio* levels of theory, and difficult at semi-empirical levels. However, molecular mechanics potentials such as TIP3P have been generated which adequately describe bulk solvent.[91] The TIP3P potential has atom centred charges in which the oxygen atom has a charge of –0.82e (and hydrogen +0.41e).[91]

Hybrid Quantum Mechanics/Molecular Mechanics[92] (QM/MM, Section 2.2.3) methods can be applied to the simulation of reactions in solution,[93,94] making such

investigations computationally amenable. With these systems, their exists a convenient partition between the QM and MM parts of the system: the solute can be treated quantum mechanically while the surrounding solvent molecules are described using the TIP3P potential. The effective Hamiltonian for the QM/MM system[14] may be given by:

$$H_{eff} = H_{QM} + H_{QM/MM} + H_{MM} + H_{bound} \tag{6.1}$$

H_{QM} is the QM Hamiltonian that incorporates the nuclear-electron attractions as well as the nuclear-nuclear and electron-electron repulsions in the QM system:

$$H_{QM} = -\frac{1}{2}\sum_{i} \nabla_i^2 - \sum_{i\alpha} \frac{Z_\alpha}{r_{i\alpha}} + \sum_{\alpha\beta} \frac{Z_\alpha Z_\beta}{r_{\alpha\beta}} + \sum_{ij} \frac{1}{r_{ij}} \tag{6.2}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

and where Z represents the charge of the QM nuclei, α and β, and r is the distance between nuclei or electrons, i and j.

The QM/MM Hamiltonian represents a cross term in which the QM part is affected by point charges of the MM system and is given by:

$$H_{QM/MM} = -\sum_{im} \frac{Q_m}{r_{im}} + \sum_{\alpha m} \frac{Z_\alpha Q_m}{r_{\alpha m}} + \sum \left\{ \frac{A_{\alpha m}}{R_{\alpha m}^{12}} - \frac{B_{\alpha m}}{R_{\alpha m}^{6}} \right\} \tag{6.3}$$

where Q_m is the charge of the MM atom, m, and r is the distance between m and the QM nuclei, α, or electrons, i. The last term in eqn. 6.3 is the Lennard-Jones 12-6 potential (L-J) that accounts for the van der Waals interaction between atoms in the MM and QM partitions; it is described by the relevant non-bonded parameters in the MM force field. The inclusion of the L-J potential allows the QM system to differentiate between like-charged, different size species (*cf.* bromide and chloride

anions); omission of this term would make any size-dependent non-bonded interactions invisible to the QM region.

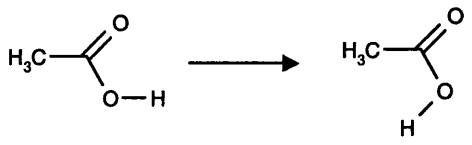
The MM Hamiltonian is dependent on atom positions and is thus represented by the MM energy.[14] The boundary Hamiltonian represents the atoms beyond the explicitly modelled system and can be split into sections that either act upon the QM or MM regions.[14] The total energy of the system is thus given by the combination of the separate QM, QM/MM, MM and boundary energies:

$$E_{tot} = E_{QM} + E_{QM/MM} + E_{MM} + E_{bound} \qquad (6.4)$$

Finally, the QM/MM interaction energy, $E_{QM/MM}$ term can be split into QM/MM electrostatic energy and van der Waals energy terms:

$$E_{QM/MM} = E_{QM/MMelec} + E_{vdw} \qquad (6.5)$$

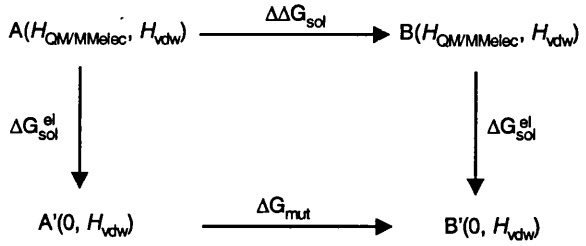
Successful QM/MM modelling using the AM1/TIP3P hybrid potential has been reported. Gao has calculated the free energy barriers for various chemical reactions[95] as well as the free energy change associated with the rotational barrier of acetic acid (scheme 6.1).[96] The latter has been calculated to be 4.6±1.2 kJ mol^{-1} as compared to the estimated experimental value of 4.2-8.4 kJ mol^{-1} in favour of the *syn* isomer.[96] Further, relative solvation free energies for many organic species are in good agreement with the experimentally determined values.[97]



syn *anti*

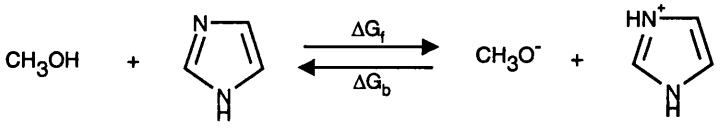
Scheme 6.1

Gao has determined these ΔG values by combining the Metropolis Monte Carlo sampling method with the hybrid QM/MM potential.[98] Briefly, the change in ΔG_{sol} between species A and B is calculated using the cycle in scheme 6.2. The free energy changes involving the gradual removal of the $H_{QM/MMElec}$ cross term are calculated for both species, leaving the MM Lennard-Jones potential as the only interaction between the MM and QM part of the system. The second stage is to calculate the free energy perturbation between A' and B' using the Monte Carlo technique to sample the conformation space. Typically, some $1.5x10^6$ configurations are used to sample this space.[98] During this part of the simulation, any major conformation changes in the solute and first solvation sphere results in the re-evaluation of the system's total energy (which includes the re-evaluation of the AM1 solute Hamiltonian).



Scheme 6.2

Bash and co-workers have used Molecular Dynamics in conjunction with the AM1/TIP3P potential in CHARMM to calculate the free energy associated with the proton transfer between methanol and imidazole.[99,100] The double width sampling employed with this work calculated the ΔG to be within *ca.* 9 kJ mol^{-1} of the experimental value.[100]



Scheme 6.3

The magnitude of the difference between the free energy values obtained for the forward process, ΔG_f and the backward process, ΔG_b gives an indication of the

sampling quality of conformation space. Thus, such a hysteresis effect can be defined as:

$$\text{Hysteresis, } H_y = |\Delta G_f - \Delta G_b| \tag{6.6}$$

A higher value for the hysteresis effect would represent poorer sampling of the conformation space. In the case of scheme 6.3,[100] the value of H_y was *ca.* 4 kJ mol^{-1}.

In calculating the free energy changes mentioned above, the following relationship is used:

$$\Delta G = -RT\ln\langle \exp(-\Delta H / RT) \rangle_A \tag{6.7}$$

where $\Delta H = H_B - H_A$ represents the change in the Hamiltonian between system A and B and $\langle\rangle_A$ is the ensemble average over a system represented by Hamiltonian, H_A.

For a very large system, Kollman has pointed out that the calculation of ΔG is more accurate than evaluating energy or enthalpy changes.[101] This is illustrated by an example that just employs a molecular mechanics potential: the transformation of methanol into ethane in which both systems are solvated by 125 water molecules.[101] The error associated with the calculated ΔG for this change is around 4 kJ mol^{-1} whereas the error associated for the corresponding enthalpy change is over ten times greater. This arises because obtaining the enthalpy change requires taking a difference in large energies, in which solvent-solvent interactions dominate. When calculating ΔG, solvent-solvent interactions remain constant (they are ensemble averaged) over the two Hamiltonia given in eqn. 6.7, leaving the solute-solvent interactions responsible for part of the energy difference. It is therefore not surprising that there are no published total energy changes between two solvated solutes while employing the QM/MM hybrid potential. However, as mentioned above, calculated free energy changes using the AM1/TIP3P potential that give good agreement with experiment have been reported.

In this chapter, a computationally efficient methodology is presented to determine the energy difference between the full sialyl α-lactone and sialyl zwitterion

intermediates in an explicit solvent environment. It also gives an opportunity to determine the reproducibility of the results obtained with the AM1/TIP3P hybrid potential, which has proven to be a successful method in calculating ΔG for various processes.

After describing the method used to calculate the energy difference between the two intermediates, the actual energy change obtained and associated error is reported and discussed. Any hysteresis effects obtained using the hybrid QM/MM method and their origins are then investigated. Comparison of energy changes between the AM1/TIP3P and AM1/COSMO methods are then discussed. This is followed by a conclusion.

6.2. Methodology

Starting structures for the sialyl α-lactone and sialyl zwitterion were initially generated with identical side chain conformations. Each intermediate was then fully optimised within a 15Å radius water droplet containing approximately 500 water molecules until the RMS gradient for the whole system had fallen to *ca.* 0.02 kJ $mol^{-1}Å^{-1}$ (0.005 kcal mol^{-1} $Å^{-1}$). The resulting solute structures are designated *la0* and *zi0* for the α-lactone and zwitterion respectively.

Conformational sampling of the solvent environment around *zi0* and *la0* was accomplished through simulated annealing techniques. These simulations provided five solvent structures around each solute. The resulting 10 structures were used as *starting points* to determine the energy difference between the two intermediates. Herein, *lass* and *ziss* refer to the α-lactone and zwitterion starting points respectively.

For each starting point, an energy profile was constructed by shortening or elongating the C_2—O_n distance in increments of 0.05Å steps (depending on whether the starting structure was a zwitterion or α-lactone). Each step on this profile was allowed 600 cycles of optimisation before the C_2—O_n distance was altered. The transition structure for each profile was refined using GRACE (section 3.3) and used

subsequently in an IRC calculation to determine the corresponding α-lactone and zwitterion structures it connects. The resulting structures were fully optimised until the RMS gradient for the whole system fell below 0.02 kJ mol^{-1} $Å^{-1}$. This process is summarised in figure 6.1, giving designations to all shown structures. Runs are labelled "forward" or "backward" depending on whether they are generated from a *lass* structure or *ziss* structure.

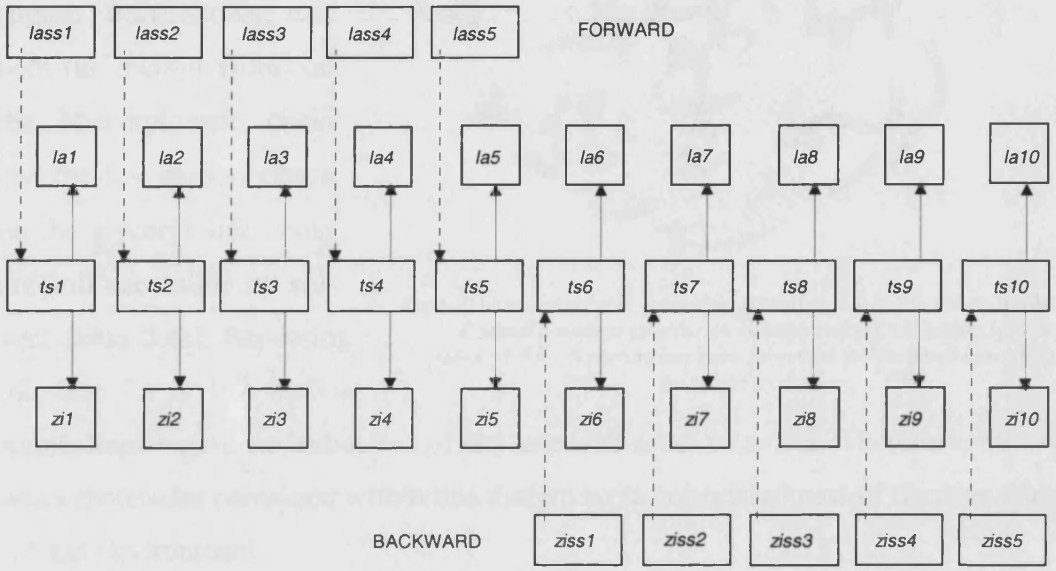


Fig 6.1. A schematic diagram showing the overall process in generating 10 pairs of intermediates, in which each pair has a different *immediate* solvent environment. Five pairs were generated from either an α-lactone starting structure (*lass*) or zwitterion starting structure (*ziss*). Dashed arrows represent profile generation as described in the text to locate the transition structure, *ts*. Continuous arrows represent the generation of α-lactone and zwitterion structures from the *ts* structure using IRC calculations followed by full QM/MM optimisation.

With the knowledge that 10 starting structures needed to be generated, applying simulated annealing to all solvent molecules in the system was considered too computationally expensive. Therefore, only the *immediate* solvent environment surrounding either *zi0* or *la0* solutes was subjected to the heating and cooling process; *bulk* solvent molecules were optimised around the annealed inner part of the system separately. Obviously, the *immediate* and *bulk* regions of the solvent needed to be adequately defined. The *immediate* solvent environment was defined as encompassing *solvent molecules that are located in at least the first solvation shell and which cover the solvent accessible area of the solute*. The remaining solvent molecules are part of the *bulk* environment.

As an initial guess, an 8Å radius water droplet was used to account for the *immediate* solvent environment around the solute. The Connolly surface[102] (figure 6.2) shows that both the methyl group on the N-acetyl side chain and the last carbon centre on the glyceryl side chain are still accessible by solvent (blue dots). Repeating this test for a 10Å radius water droplet gave no indication of any exposed areas of solute. Therefore, all 140 water molecules contained within this system were considered part of the *immediate* solvent environment.

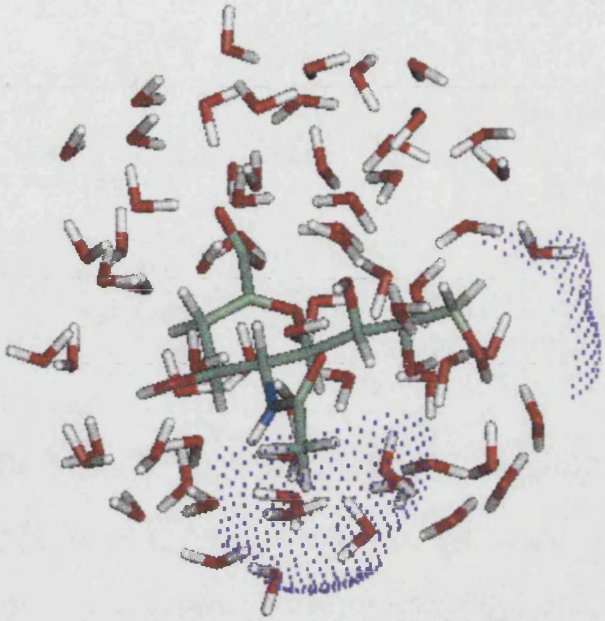


Fig 6.2. Exposed solvent accessible area of solute in 8Å water droplet using Connolly surface calculation as implemented in Quanta (probe radius=1.4Å). Structure has been subjected to simulated annealing as described in the text.

To generate the 10 starting points, both pyranose-oxygen centred solutes were placed in the 10Å radius water droplet and fixed throughout the simulated annealing procedure. In addition, a 10Å stochastic boundary was applied to prevent any evaporation of solvent molecules from occurring. Prior to the heating phase, the solvent molecules were optimised using MM only, until the gradient fell below 0.2 kJ mol^{-1} $Å^{-1}$ (0.05 kcal mol^{-1} $Å^{-1}$). The system (described with an MM only potential) was then rapidly heated to 300K from absolute zero over 6ps, equilibrated for a further 34ps and subsequently cooled back down to absolute zero over 160ps. The resulting solvent structure was then MM minimised until the RMS gradient fell below 0.02 kJ $mol^{-1}Å^{-1}$. Test runs suggested that longer cooling periods in the simulated annealing process were not sufficiently advantageous to warrant their use (a 280ps cooling period resulted in a structure that was lower in energy by only 4 kJ mol^{-1}). The resulting structure was then subjected to further simulated annealing (using the same method as mentioned above) to give a new *immediate* solvent

environment. This process was continued to give five different solvent environments around each intermediate.

Each solute-*immediate* solvent structure was then placed into 15Å radius *bulk* solvent sphere (in which there exists a central cavity). The *immediate* solvent region and solute were fixed while optimisation of the *bulk* solvent was carried out in the presence of the QM described solute. The solute and *immediate* solvent were then released and the whole system was fully optimised using the AM1/TIP3P hybrid potential until the RMS gradient was below 0.02 kJ $mol^{-1}Å^{-1}$. The process of setting up the starting structures is summarised in figure 6.3.

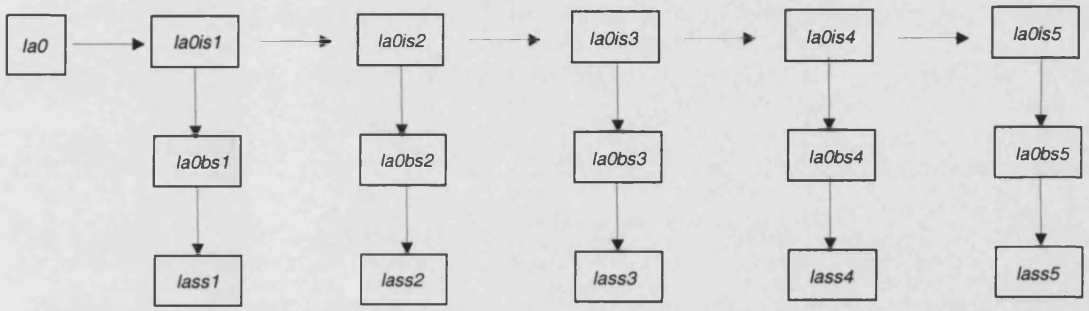


Fig 6.3. Generation of α-lactone starting structures. *la0* is solvated in the 10Å *immediate* solvent sphere and the resulting system is subjected to simulated annealing as prescribed in the text. The resulting α-lactone–*immediate* solvent structure (*la0is*) is then subjected to more simulated annealing to generate the next *la0is* structure. This procedure is repeated to give five different solvent environments. Each *la0is* structure is placed into the *bulk* solvent shell and is fixed while the *bulk* solvent molecules optimise around it. This process gives the *la0bs* structures. Finally, the inner system (*la0is*) is released and the whole system is fully optimised to give the α-lactone starting structures (*lass*). The ziss structures were generated in precisely the same way.

6.3. α-Lactone-Zwitterion Energy Difference

Table 6.1 lists the total energies for the ten α-lactone and zwitterion intermediates, along with the corresponding energy changes. In every single case, the zwitterion is lower in energy than the α-lactone. The average energy difference between both intermediates is found to be –60.1 ± 32.6 kJ mol^{-1}. The value for the standard deviation (32.6 kJ mol^{-1}) is in agreement with the error associated with calculated energy changes stipulated by Kollman.[101] In this case, the smallest energy difference

between both intermediates is –5.2 kJ mol^{-1} (la3→zi3), while the largest difference is -102.7 kJ mol^{-1} (la9→zi9).

Table 6.1. Total energies for α-lactone and zwitterion intermediates obtained for all runs, along with the corresponding energy difference between both intermediates. Averages and standard deviations are also given over runs 1-10, runs 1-5 and runs 6-10

Run	α-Lactone (la) Energy/ kJ mol^{-1}	Zwitterion(zi) Energy/ kJ mol^{-1}	α-Lactone→Zwitterion / kJ mol^{-1}
1	-26564.26	-26602.00	-37.7
2	-26589.70	-26609.53	-19.8
3	-26432.63	-26437.82	-5.2
4	-26438.44	-26485.89	-47.4
5	-26580.87	-26623.21	-42.3
6	-26423.97	-26517.31	-93.3
7	-26569.95	-26638.98	-69.0
8	-26455.47	-26552.42	-96.9
9	-26613.50	-26716.18	-102.7
10	-26486.77	-26572.92	-86.1
Average (1-10)	**-26515.6**	**-26575.6**	**-60.1**
Std. Deviation	**70.9**	**76.7**	**32.6**
Average (1-5)	**-26521.2**	**-26551.7**	**-30.5**
Std. Deviation	**78.7**	**84.1**	**17.6**
Average (6-10)	**-26509.9**	**-26599.6**	**-89.6**
Std. Deviation	**79.4**	**78.8**	**13.0**
Hysteresis			**59.1**

From table 6.1, it is evident that a "hysteresis" effect exists, which is analogous to that given in eqn. 6.6 for the free energy. In this study, the magnitude of the α-lactone-zwitterion energy is dependent on whether the starting structure is an α-lactone (runs 1-5) or zwitterion (runs 6-10). This would initially suggest, especially for the charge-separated zwitterion, that the immediate solvent structure plays a

crucial role in stabilising the solute. The hysteresis effect may indicate that a solvent structure has a bias for stabilising one intermediate at the expense of the other. In this case, the α-lactone-zwitterion energy difference is -30.5 ± 17.5 kJ mol^{-1} when initially starting from an α-lactone intermediate and -89.6 ± 13.0 kJ mol^{-1} when starting from the zwitterion. The resulting hysteresis effect here is large with a value of 59.1 kJ mol^{-1}. It would be therefore worthwhile determining the origins of this effect.

6.4. Hysteresis and Solvent Effects

To determine the origin of the large hysteresis effect, average energy changes between both intermediates for runs (1-5) and runs (6-10) were evaluated in terms of solute, *bulk* solvent, *immediate* solvent and the interaction energy between solute and solvent. This assumes that the total energy of system can be given as:

$$E_{tot} = E_{sol} + E_{solint} + E_{imm} + E_{bulk} \tag{6.8}$$

where, E_{sol} is the energy of the solute, E_{solint} is the solute-solvent interaction energy, E_{imm} is the *immediate* solvent energy (solvent-solvent interactions) and E_{bulk} is the *bulk* solvent energy. The interaction between the *immediate* and *bulk* solvent is included in the E_{bulk} term. The solute energy, E_{sol} is given by:

$$E_{sol} = \langle \psi | H_{QM} | \psi \rangle \tag{6.9}$$

where, H_{QM} is the solute Hamiltonian given in eqn 6.2.

E_{solint} is equivalent to the QM/MM interaction energy term given eqn. 6.5 and is thus calculated using:

$$E_{solint} = \langle \psi | H_{QM} + H_{QM/MM} | \psi \rangle - \langle \psi^o | H^o_{QM} | \psi^o \rangle \tag{6.10}$$

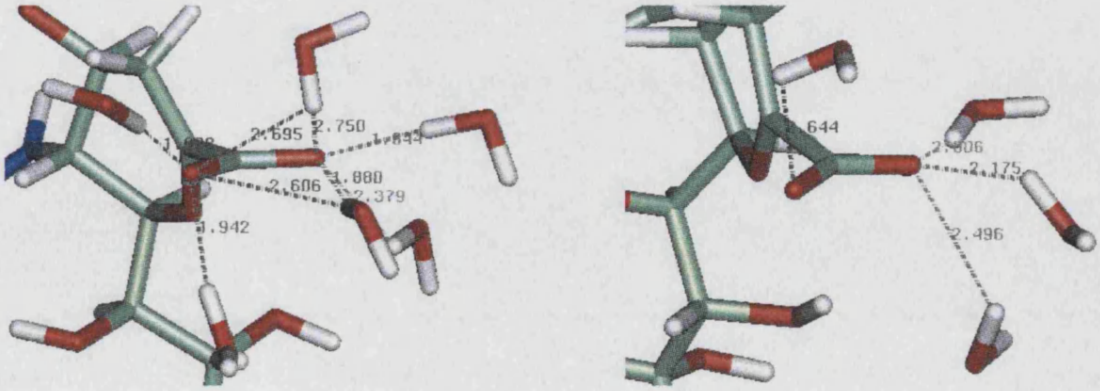
where H_{QM} and $H_{QM/MM}$ are defined in equations 6.2 and 6.3 respectively and H^{o} is the Hamiltonian for the solute structure in the gas phase (eqn. 6.2). E_{solint} is also divided into contributions from *bulk* and *immediate* solvent:

$$E_{solint} = E_{intimm} + E_{intbulk} \tag{6.11}$$

Finally, the sum of the solvent energies, E_{bulk} and E_{imm} is equal to the MM energy:

$$E_{mm} = E_{bulk} + E_{imm} \tag{6.12}$$

The energy difference between the α-lactone and zwitterion is given in terms of these energy contributions in table 6.2. From these results, it is evident that the solute makes a moderate contribution (10.7 kJ mol^{-1}) to H_y, while the solvent contributes negligibly to this effect (4.3 kJ mol^{-1}). However, this small value results from a cancellation of higher magnitude H_y effects associated with the *immediate* (14.5 kJ mol^{-1}) and *bulk* (-10.2 kJ mol^{-1}) solvent. The most significant contribution to the overall hysteresis effect comes from the solute-solvent interaction terms of which the majority is associated with the solute-*immediate* solvent interactions (48.3 kJ mol^{-1}). The observed hysteresis effect suggests the *immediate* solvent structure preferentially stabilises the initial intermediate compared to the corresponding counter solute. Evidence for this effect can also be seen visually when considering the solvent structure around the carboxylate centre of the zwitterion.



a). b).

Fig 6.4. Close interactions (<2.8Å) between the solute carboxylate group and *immediate* solvent molecules a). from a zwitterion starting point (*ziss9*) and b). from an α-lactone starting point (*lass3*).

Table 6.2. Energy difference between α-lactone and zwitterion in terms of the solute, ΔE_{sol}, and solute interaction energy with solvent, ΔE_{solint}, which comprises of the solute interaction with immediate, ΔE_{intimm}, and bulk solvent, $\Delta E_{intbulk}$. Further contributions also occur from the solvent energy, ΔE_{mm}, which comprise of *immediate* solvent, ΔE_{imm}, and *bulk* solvent, ΔE_{bulk}.

Run	ΔE_{sol} /kJ mol^{-1}	ΔE_{intimm} /kJ mol^{-1}	$\Delta E_{intbulk}$ /kJ mol^{-1}	ΔE_{solint} /kJ mol^{-1}	ΔE_{imm} /kJ mol^{-1}	ΔE_{bulk} /kJ mol^{-1}	ΔE_{mm} /kJ mol^{-1}	$\Delta E_{sol}+\Delta E_{solint}$ /kJ mol^{-1}
1	-0.98	-92.63	-12.47	-105.10	59.73	8.60	68.34	-106.08
2	0.22	-54.43	1.87	-52.57	33.36	-0.85	32.51	-52.35
3	0.58	-47.30	5.06	-42.24	33.21	3.26	36.47	-41.66
4	-2.81	-81.39	-0.29	-81.68	85.02	-47.97	37.04	-84.49
5	-6.40	-65.97	-4.79	-70.76	42.57	-7.74	34.82	-77.17
6	-10.07	-129.38	5.17	-124.21	35.86	5.08	40.94	-134.29
7	-6.33	-97.16	8.72	-88.44	42.08	-16.34	25.74	-94.78
8	-16.03	-119.44	2.34	-117.10	37.98	-1.79	36.19	-133.13
9	-16.66	-140.48	-5.91	-146.39	43.96	16.42	60.38	-163.05
10	-13.79	-96.95	0.18	-96.76	21.38	3.03	24.41	-110.56
Average (1-5)	**-1.9**	**-68.3**	**-2.1**	**-70.5**	**50.8**	**-8.94**	**41.8**	**-72.4**
Average(6-10)	**-12.6**	**-116.7**	**2.1**	**-114.6**	**36.3**	**1.3**	**37.5**	**-127.2**
Hysteresis	**10.7**	**48.3**	**-4.2**	**44.1**	**14.5**	**-10.2**	**4.3**	**54.8**

Table 6.3. α-Lactone energies in terms of the solute, E_{sol}, and solute interaction energy with solvent, E_{solint}, which comprises of the solute interaction with *immediate*, E_{intimm}, and *bulk*, $E_{intbulk}$, solvent. Further contributions also occur from the solvent energy, E_{mm}, which comprise of *immediate* solvent, E_{imm}, and *bulk* solvent, E_{bulk}.

Run	E_{sol} /kJ mol^{-1}	E_{intimm} /kJ mol^{-1}	$E_{intbulk}$ /kJ mol^{-1}	E_{solint} /kJ mol^{-1}	E_{imm} /kJ mol^{-1}	E_{bulk} /kJ mol^{-1}	E_{mm} /kJ mol^{-1}
1	-1264.09	-268.93	-4.27	-273.20	-6604.87	-18422.10	-25026.97
2	-1270.35	-273.33	-18.64	-291.97	-6576.79	-18450.59	-25027.38
3	-1275.49	-248.78	-11.00	-259.78	-6732.94	-18164.42	-24897.36
4	-1275.26	-265.11	-14.94	-280.05	-6680.31	-18202.82	-24883.14
5	-1263.82	-283.99	-19.32	-303.31	-6530.48	-18483.25	-25013.73
6	-1259.53	-259.98	-13.72	-273.69	-6555.17	-18335.58	-24890.75
7	-1256.15	-289.28	-22.15	-311.42	-6573.92	-18428.46	-25002.37
8	-1252.24	-324.83	-19.79	-344.62	-6431.66	-18426.95	-24858.61
9	-1250.14	-289.79	-24.12	-313.91	-6531.38	-18518.07	-25049.45
10	-1251.53	-302.36	-19.64	-322.00	-6505.15	-18408.09	-24913.24

Table 6.4. Zwitterion energies in terms of the solute, E_{sol}, and solute interaction energy with solvent, E_{solint}, which comprises of the solute interaction with *immediate*, E_{intimm}, and *bulk*, $E_{intbulk}$, solvent. Further contributions also occur from the solvent energy, E_{mm}, which comprise of *immediate* solvent, E_{imm}, and *bulk* solvent, E_{bulk}.

Run	E_{sol} /kJ mol^{-1}	E_{intimm} /kJ mol^{-1}	$E_{intbulk}$ /kJ mol^{-1}	E_{solint} /kJ mol^{-1}	E_{imm} /kJ mol^{-1}	E_{bulk} /kJ mol^{-1}	E_{mm} /kJ mol^{-1}
1	-1265.07	-361.56	-16.74	-378.30	-6545.13	-18413.50	-24958.63
2	-1270.13	-327.77	-16.77	-344.54	-6543.43	-18451.43	-24994.86
3	-1274.91	-296.08	-5.93	-302.02	-6699.73	-18161.16	-24860.89
4	-1278.07	-346.50	-15.23	-361.73	-6595.29	-18288.45	-24883.75
5	-1270.22	-349.97	-24.11	-374.08	-6487.91	-18491.00	-24978.91
6	-1269.60	-389.36	-8.55	-397.91	-6519.32	-18330.49	-24849.81
7	-1262.48	-386.43	-13.43	-399.86	-6531.84	-18444.80	-24976.63
8	-1268.27	-444.27	-17.45	-461.73	-6393.68	-18428.75	-24822.43
9	-1266.80	-430.27	-30.03	-460.30	-6487.42	-18501.66	-24989.07
10	-1265.33	-399.31	-19.45	-418.76	-6483.77	-18405.06	-24888.83

As can be seen in figure 6.4, the solvent structure surrounding the zwitterion carboxylate group (obtained from a *ziss* species) has some eight interactions with neighbouring solvent molecules. However, the resulting zwitterion obtained from an α-lactone starting structure (*lass*) leads to a smaller number of strong hydrogen bonding interactions between solvent and the carboxylate group. The corresponding difference in solvent interactions for the α-lactone intermediate is far less obvious than for the charged separated zwitterion.

These differing solvent environments also effect the solute structure. In the α-lactone, the C_2—O_n bond is on average 0.017Å longer (Table 6.5) in structures la6-la10 (starting from ziss species). This difference is also accompanied by a larger C_2—C_1—O_n bond angle (by about 1°) and a shorter C_2—O_h bond (0.009Å), although there is no change in the planarity around the C_2 atom (ΣC_2). However, the most significant change is the C_6—O_h—C_2—C_3 pyranose-ring torsion angle, which is smaller by 27.3° in α-lactones *la6-la10* compared to structures *la1-la5*. This change is visibly seen in fig 6.5. The effect of the radically different solvent environment when starting from a *ziss* species results in the flattening of the $^2C_5$ conformation of the pyranose ring. However, this effect is exaggerated because one spiro-α-lactone adopts a $B_{2,5}$ conformation (*la9*).

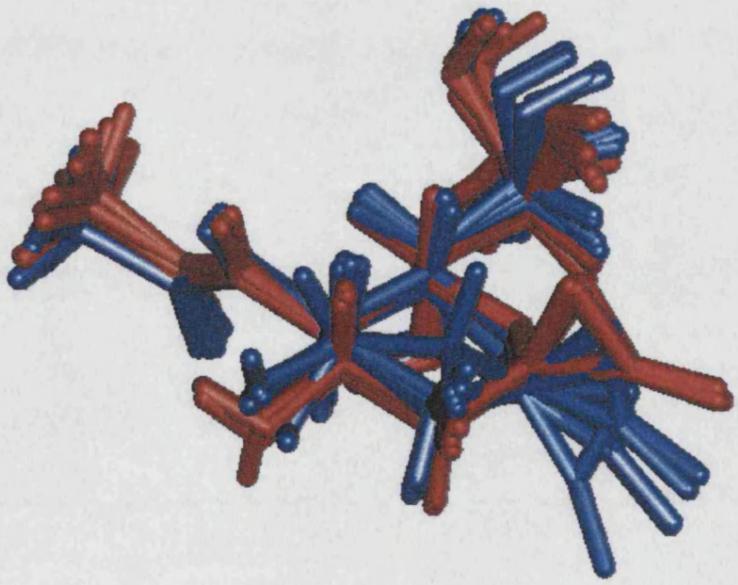
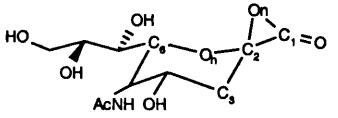


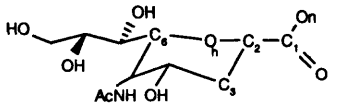
Fig 6.5. Comparison of all 10 α-lactone structures, generated from runs that use either an α-lactone starting point (red) or a zwitterion starting point (blue).

Table 6.5. Selected geometrical parameters (bonds in Ångstroms, angles in degrees) for the sialyl α-lactone intermediate.



Run	C_2-O_h	C_2-O_n	ΣC_2	O_n-C_2-C_1	O_h-C_2-C_1-O_n	C_6-O_h-C_2-C_3	O_h-C_3-C_2-C_1
1	1.384	1.491	357.4	63.3	97.9	47.1	199.4
2	1.385	1.491	357.3	63.3	96.5	48.8	199.9
3	1.385	1.490	357.4	63.3	96.6	52.4	199.1
4	1.383	1.493	357.6	63.6	97.7	51.5	198.4
5	1.387	1.485	357.8	63.2	97.1	58.0	198.0
6	1.376	1.510	356.9	64.6	96.7	36.2	201.3
7	1.380	1.509	357.5	64.5	95.1	44.8	198.7
8	1.373	1.516	357.9	64.7	95.8	16.9	198.0
9	1.382	1.492	358.1	63.4	94.4	-7.2	196.3
10	1.378	1.507	357.5	64.3	93.8	30.5	199.2
Average(1-5)	**1.385**	**1.490**	**357.5**	**63.3**	**97.2**	**51.6**	**199.0**
Std. Deviation	**0.001**	**0.003**	**0.2**	**0.2**	**0.6**	**4.2**	**0.8**
Average(6-10)	**1.378**	**1.507**	**357.6**	**64.3**	**95.2**	**24.2**	**198.7**
Std. Deviation	**0.003**	**0.009**	**0.5**	**0.5**	**1.1**	**20.3**	**1.8**
Difference	**-0.007**	**0.017**	**0.1**	**1.0**	**-2.0**	**-27.3**	**-0.3**

Table 6.6. Selected geometrical parameters (bonds in Ångstroms, angles in degrees) for the sialyl zwitterion intermediate.



Run	C_2-O_h	C_2-O_n	ΣC_2	O_n-C_2-C_1	O_h-C_2-C_1-O_n	C_6-O_h-C_2-C_3	O_h-C_3-C_2-C_1
1	1.294	2.327	359.6	110.0	88.8	6.6	172.6
2	1.296	2.328	359.6	109.1	61.4	6.8	171.7
3	1.296	2.274	358.5	105.5	73.7	11.7	164.9
4	1.294	2.338	359.6	111.4	80.5	6.1	173.4
5	1.294	2.380	360.0	114.9	103.7	1.4	181.5
6	1.291	2.387	360.0	116.1	74.1	3.0	181.8
7	1.293	2.375	359.7	114.7	52.3	4.0	177.6
8	1.294	2.358	359.9	113.8	79.4	4.7	176.3
9	1.293	2.382	360.1	116.5	74.6	0.9	180.7
10	1.295	2.398	360.0	116.7	57.8	0.9	181.6
Average(1-5)	**1.295**	**2.329**	**359.5**	**110.2**	**81.6**	**6.5**	**172.8**
Std. Deviation	**0.001**	**0.038**	**0.6**	**3.4**	**15.9**	**3.7**	**5.9**
Average(6-10)	**1.293**	**2.380**	**359.9**	**115.6**	**67.6**	**2.7**	**179.6**
Std. Deviation	**0.001**	**0.015**	**0.2**	**1.3**	**11.8**	**1.8**	**2.5**
Difference	**-0.002**	**0.051**	**0.5**	**5.4**	**-14.0**	**-3.8**	**6.8**

Changes in the corresponding zwitterion structures are also evident. The C_2—O_n distance is 0.051Å shorter for zwitterions *zi1-zi5* (generated in the runs using the *lass* species). There is also a concomitant decrease in the C_2—C_1—O_n angle of, on average, 5.4°. The carboxylate group is also slightly elevated from the pyranose ring-plane as evidenced by a smaller average O_h—C_3—C_2—C_1 dihedral angle (by some 6.8°). The comparison of all 10 zwitterion structures is shown in figure 6.6.

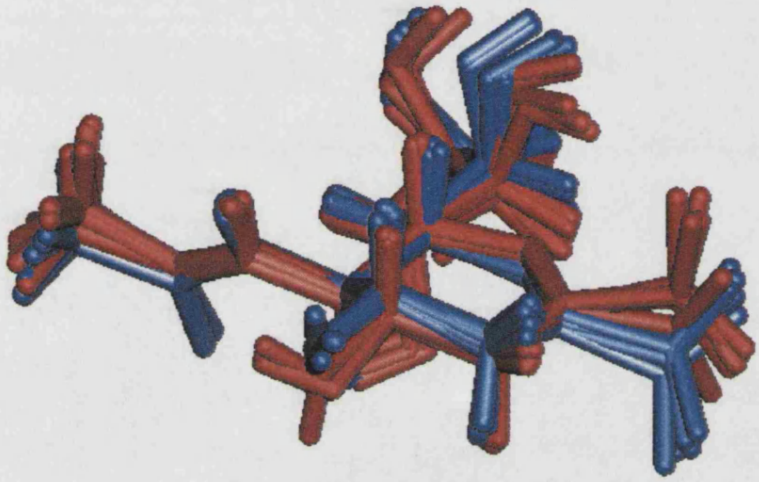
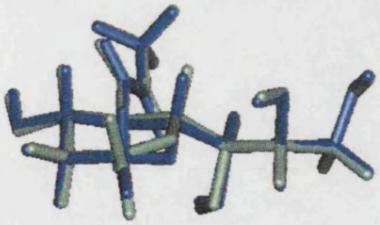


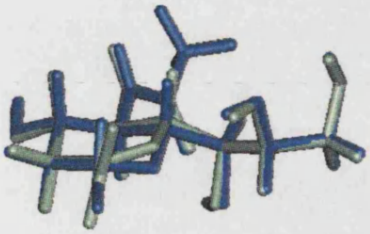
Fig 6.6. Comparison of the 10 zwitterion structures, generated from runs that use either an α-lactone (red) or zwitterion (blue) as starting points.

Figure 6.7 provides evidence that the α-lactone-zwitterion transition involves a consistent structural change in moving from one intermediate to the other. In all but two runs, the RMS deviation between the zwitterion and α-lactone structure is between 0.19Å and 0.31Å. In run 9, a large RMS deviation is obtained because the α-lactone adopts a $B_{2,5}$ conformation rather than a $^2C_5$ conformation. In run 5, the pyranose ring dihedral angle C_6—O_h—C_2—C_3 in the α-lactone deviates away from its corresponding value in the zwitterion by about 57°. This is the greatest change in that torsion angle between the two intermediates for any of the studied runs. Elimination of runs 9 and 5 to the evaluation of the α-lactone-zwitterion energy difference has a negligible effect on the hysteresis value, while the average energy change is slightly lower at 57.0 ± 32.5 kJ mol^{-1}.

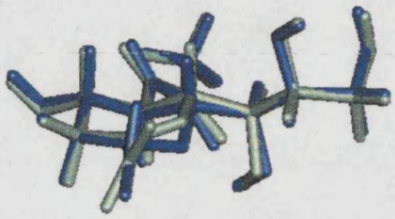


RUN 1: RMS=0.26Å

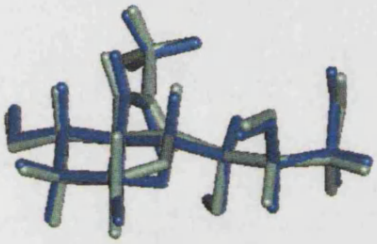
RUN 6: RMS=0.28Å



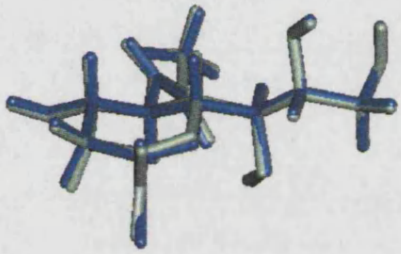
RUN 2: RMS=0.26Å



RUN 7: RMS=0.31Å

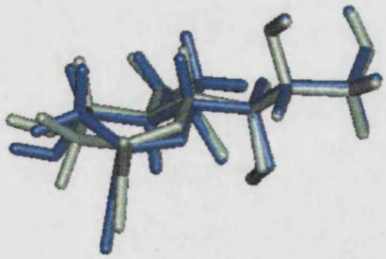


RUN 3: RMS=0.25Å



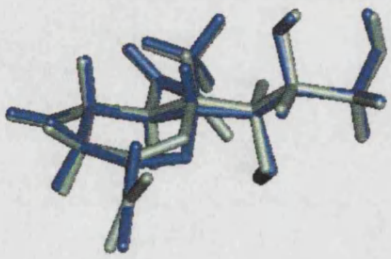
RUN 8: RMS=0.19Å

RUN 4: RMS=0.31Å



RUN 9: RMS=0.49Å

RUN 5: RMS=0.47Å



RUN 10: RMS=0.27Å

Fig 6.7. Comparison of α-lactone (blue) and zwitterion (green) intermediates for all runs.

Figure 6.7 also shows how sensitive the zwitterion carboxylate group is to the solvent environment, giving large variations in the O_h—C_2—C_1—O_n dihedral angles which is accompanied by large values for the corresponding standard deviation (Table 6.6).

In comparing the structures obtained from either the forward and backward processes (figures 6.5 and 6.6), it is evident that there are different side chain conformations involving the 4-OH and 8-OH moieties. This points evidence towards different conformations in the starting solutes, *la0* and *zi0*. Indeed, comparison of these two systems (fig 6.8) confirms this. Such changes in the *zi0* and *la0* structures were generated within the 15Å water droplet (prior to this optimisation, both species had identical side chain conformations). These changes in side-chain conformations have probably enhanced the hysteresis effect associated with the solute.

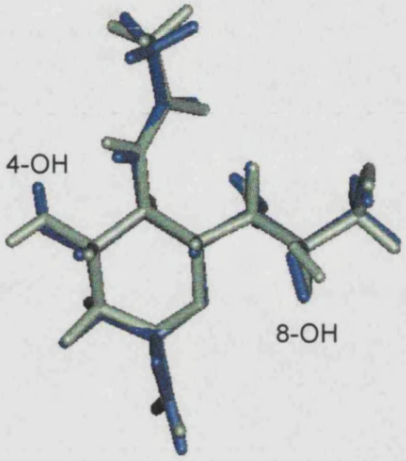


Fig 6.8. Comparison of *la0* (blue) and *zi0* (green) initial solute structures.

6.5. Transition Structures

Although the previous section considered the solvent environment and its effects on the intermediates, these structures are derived from the connecting transition structure for the opening/closure of the α-lactone ring. Thus, any changes in the solvent environment may effect the transition structure and consequently the resulting reactant and product structures obtained from it. Table 6.7 gives some selected geometrical parameters for the 10 transition structures. For runs starting from a zwitterion species (backward) the average C—O_n bond length of 1.697Å is around 0.107Å shorter than the average C—O_n length for runs that start from an α-lactone (forward). This difference is accompanied with a longer C_2—O_h bond, giving values of 1.340Å and 1.349Å for the forward and backward runs respectively. Further, the C_2—C_1—O_n angle has a smaller average value for the backward runs (74.7°) compared to the average value obtained for the forward runs (79.9°). The

carboxylate group is also further out of the pyranose plane: for the backward transition structures, the O_h—C_3—C_2—C_1 dihedral has a smaller (171.6°) value than for the "forward" transition structures (182.9°). These average differences between the two sets of transition structures can be seen in figure 6.9.

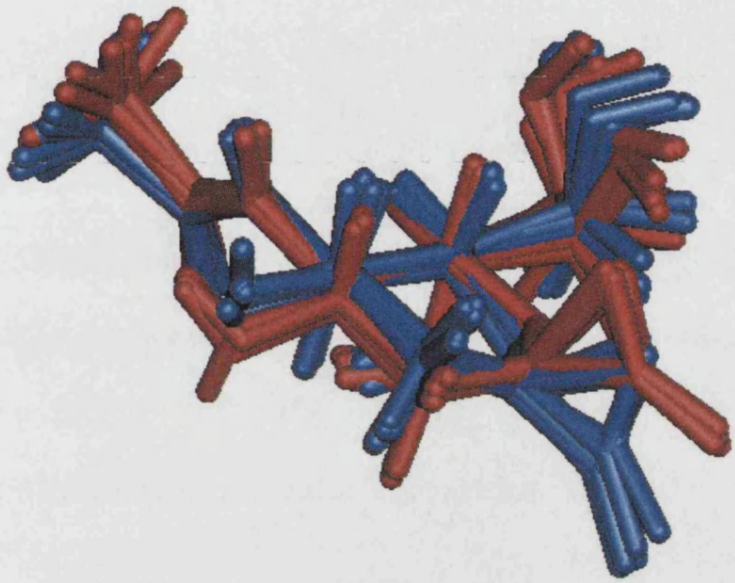
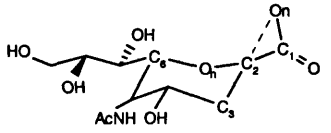


Fig 6.9. Transition structures for the α-lactone ring opening/closure. Five TS structures are generated from α-lactone starting structures (red) and five are generated from zwitterion starting structures (blue).

Figure 6.9 shows the deviation of the forward and backward transition structures. Moliner *et al*, in the modelling of the lactate dehydrogenase catalysed conversion of pyruvate to lactate, have suggested that the essential elements of the TS structure are robust with respect to its environment.[103] For a particular chemical change under study, Moliner *et al.* showed that a flexible environment is able to adopt many different conformations at the first order saddle point.[103] This led to a suggestion that for large systems, a family of transition structures exists.[103] In the simulations presented here, there is invariance in the TS structures obtained from all runs in the same direction, but an obvious discrepancy in the TS structures obtained from runs in different directions. One of the major differences is in the reaction co-ordinate itself: the C_2—O_n bond length is on average 0.107Å longer in the "forward TS structures" than in the "backward TS structures". The runs that start from α-lactones, tend to lead to TS structures with an increased degree of zwitterionic character (i.e. they are more product-like). Furthermore, the corresponding runs that start from

Table 6.7. Selected geometrical parameters (bonds in Ångstroms, angles in degrees) for the transition structures.



Run	C_2-O_h	C_2-O_n	ΣC_2	C_2-C_1-O_n	O_h-C_2-C_1-O_n	C_6-O_h-C_2-C_3	O_h-C_3-C_2-C_1
1	1.341	1.788	359.7	79.0	88.8	28.9	173.3
2	1.342	1.796	359.6	79.4	86.5	31.8	172.6
3	1.340	1.815	359.4	80.4	85.6	35.2	170.4
4	1.341	1.794	359.6	79.6	88.5	33.3	172.4
5	1.338	1.828	359.2	81.2	87.8	36.6	169.1
6	1.345	1.700	359.8	74.9	89.4	15.0	184.5
7	1.345	1.724	360.0	76.1	86.9	28.7	180.3
8	1.352	1.680	359.9	73.7	91.4	11.7	183.8
9	1.353	1.665	359.8	73.1	91.2	11.5	185.3
10	1.350	1.718	360.0	75.9	85.7	18.1	180.8
Average(1-5)	**1.340**	**1.804**	**359.5**	**79.9**	**87.4**	**33.2**	**171.6**
Std. Deviation	**0.002**	**0.017**	**0.2**	**0.9**	**1.4**	**3.0**	**1.7**
Average(6-10)	**1.349**	**1.697**	**359.9**	**74.7**	**88.9**	**17.0**	**182.9**
Std. Deviation	**0.004**	**0.025**	**0.1**	**1.3**	**2.6**	**7.1**	**2.3**
Difference	**0.009**	**-0.107**	**0.4**	**-5.2**	**1.5**	**-16.2**	**11.4**

zwitterion intermediates tend to yield TS structures with more α-lactone like character. This suggests that *immediate* solvent structures, which are optimised to solvate a particular solute, influence the resulting TS structure in agreement with the Hammond postulate. This could suggest that in simulating reactions in solution, the TS structure is dependent on whether the simulation is started from the reactant or product wells. In this case, the change may be emphasised somewhat because one intermediate possesses separated charges (but is neutral overall) while the other does not.

6.6. AM1/COSMO Energies of Both Intermediates in Runs 1-10

Table 6.8. gives the energies of both intermediates calculated using the AM1/COSMO continuum solvation method for all runs, with the corresponding α-lactone-zwitterion energy difference.

The average energy change for conversion of α-lactone to zwitterion is calculated to be -100.4 ± 13.9 kJ mol^{-1} using the AM1/COSMO method. The energy change is thus 40 kJ mol^{-1} larger than that obtained using the AM1/TIP3P method, although it is just below the lower bounds of the QM/MM work (-92.7 kJ mol^{-1}). Therefore, this suggests that AM1/COSMO seems to over estimate energy changes between the two intermediates compared to the AM1/TIP3P result.

The hysteresis effect for the AM1/COSMO energies is calculated as 20.0 kJ mol^{-1}. As mentioned above, the contribution by the (*in vacuo*) solute to this effect is 10.7 kJ mol^{-1}; the remainder is due to the interaction between the solute and the continuum (9.3 kJ mol^{-1}). This suggests the interaction of the AM1/COSMO continuum environment with a solute structure is less sensitive as compared to using an AM1/TIP3P method.

Table 6.8. AM1/COSMO energies for the α-lactone and zwitterion intermediates in all 10 runs, along with the corresponding energy difference.

Starting Structure	Lactone Energy /kJ mol^{-1}	Zwitterion Energy /kJ mol^{-1}	ΔE /kJ mol^{-1}
1	-1410.04	-1501.73	-91.69
2	-1411.58	-1495.01	-83.43
3	-1415.52	-1489.13	-73.61
4	-1416.66	-1511.93	-95.27
5	-1409.07	-1516.87	-107.80
6	-1412.39	-1520.67	-108.28
7	-1410.22	-1514.27	-104.05
8	-1407.70	-1520.73	-113.03
9	-1410.28	-1523.69	-113.41
10	-1404.95	-1518.02	-113.07
Average (1-10)	**-1410.8**	**-1511.2**	**-100.4**
Std. Deviation	**3.5**	**11.9**	**13.9**
Average (1-5)	**-1412.6**	**-1502.9**	**-90.4**
Std. Deviation	**3.4**	**11.5**	**12.8**
Average (6-10)	**-1409.1**	**-1519.5**	**-110.4**
Std. Deviation	**2.9**	**3.5**	**4.1**
Hysteresis			**20.0**

6.7. Conclusions

A methodology has been designed to evaluate the AM1/TIP3P energy difference between two chemical species in solution and the approximate error associated with it. The method essentially divides the solvent into two regions, in which the *immediate* solvent essentially covers the solvent accessible area of the solute. This inner solvent partition is subjected to numerous heating and cooling cycles to provide alternative solvent structures around the solute. The outer sphere of *bulk* solvent is then added and the whole system is subsequently optimised. Both species under study are used as starting points and a reaction profile is generated from them

to obtain the TS which connects both structures. This is then subsequently used in an IRC calculation to verify the reactant and product structures.

The energy difference between the α-lactone and zwitterion was calculated as -60.1 ± 32.6 kJ mol^{-1}, where the zwitterion is the lower energy species in solution for all 10 runs. However, the resulting hysteresis effect of about 59 kJ mol^{-1} is associated with the different directions taken along the reaction co-ordinate. The biggest contribution to this effect is the solute-*immediate* solvent interaction and this suggests the *immediate* solvent will preferentially stabilise the intermediate originally contained within it. The solute structure is also different depending on the direction taken along the reaction co-ordinate, but relative changes between α-lactone and zwitterion species seem consistent for most of the runs generated here. In contrast to the AM1/TIP3P energy, the corresponding change between both intermediates at AM1/COSMO is calculated as -100.4 ± 13.9 kJ mol^{-1}, which is somewhat larger. It is also noted that the continuum method is less "responsive" to the solute structure compared to explicit solvation methods.

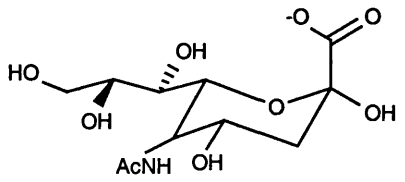
The structural differences obtained for each intermediate in both the forward and backward directions may originate from the slightly different TS structures obtained for both processes. The TS may therefore be affected by major changes in the immediate environment, while smaller changes will lead to essentially the same structure. Modelling chemical reactions in solution using explicit solvation methods may therefore require the investigation of both the forward (reactants to products) and reverse (products to reactants) processes to determine the TS structure for a chemical change. In reality, such processes should be micro-reversible; however, the simulation of chemical reactivity, as implemented in this chapter, is strongly dependent on the *immediate* solvent structure.

7. STRUCTURE, FUNCTION AND INHIBITION OF INFLUENZA SIALIDASE

7.1. The Influenza Virus and its Life Cycle

The Orthomyxovirus, influenza, infects several species in the animal kingdom, including birds, pigs, horses, whales, seals, mink and humans.[104] Three types of influenza virus are known, A, B and C, each distinguished by a specific set of RNA codes. Influenza B and influenza C only infect humans, but influenza A also infects other species and is responsible for numerous morbidity and mortality cases each year. One frightening example of influenza's virulence is the "Spanish influenza" pandemic of 1918 which resulted in the death of at least 20 million humans.[104]

The *nucleocapsid* of the enveloped influenza-A virion (figure 7.1) consists of eight strands of RNA enclosed within the matrix protein, M_1.[105-107] Surrounding the nucleocapsid is a lipid bi-layer, above which sit two glycoproteins, *hemagglutinin* (HA) and *neuraminidase* (NA, *sialidase*, acylneuraminyl hydrolase, E.C. 3.2.1.18). HA molecules are rod-shaped trimers that cover 90% of the virion surface, while NA species are mushroom-shaped tetramers which cover the remaining 10% of the particle. Both glycoproteins have a long stalk that penetrates the lipid bi-layer and terminates with a knob that anchors them to the virus membrane.[104] Both antigens recognise a terminal sialic acid receptor related to N-acetyl-α-D-neuraminic acid (Neu5ac, **1**); their importance is only realised when considering the life cycle of the influenza virus (figure 7.2).[104-108]



1

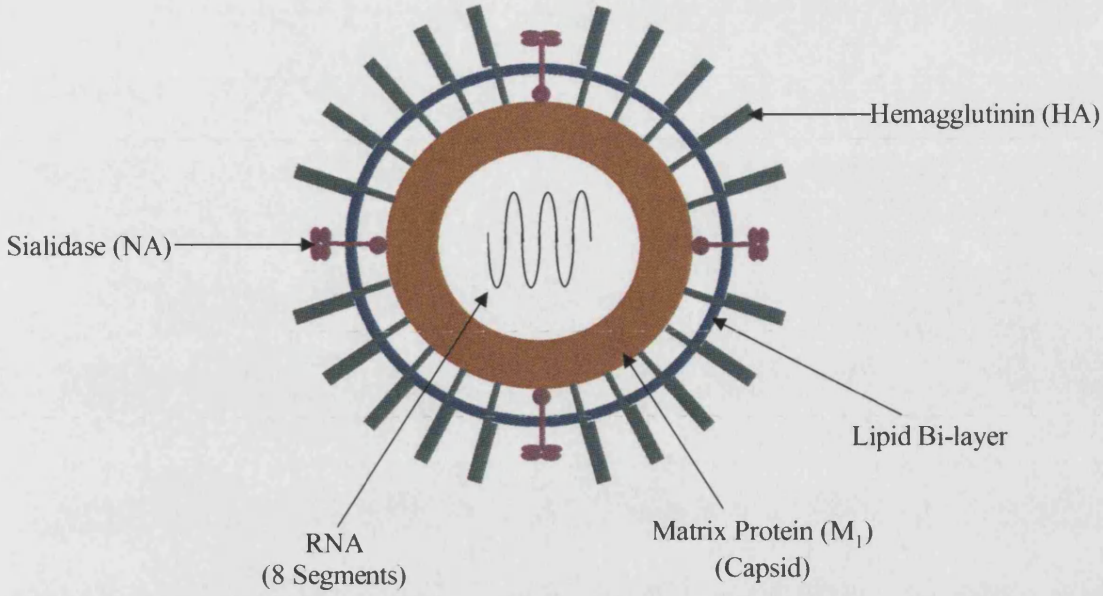


Fig 7.1. Basic schematic representation of an influenza virion

Influenza infection is usually restricted to epithelial cells located in the respiratory tract of the host. HA initiates the viral infection by attachment to the Neu5ac receptor, found on the surface glycoprotein of the host cell (Fig 7.2, stage 1). The virus is then absorbed into the cell (*receptor-mediated endocytosis*, Fig 7.2, stage 2), which eventually results in the fusion of both viral and host cell membranes. Two processes simultaneously ensue in the replication of new virions. Firstly, a relatively fast stage involves the generation of new viral glycoproteins (HA and NA) which accumulate on the host cell surface (Fig 7.2, stages 4&5). Secondly, a much slower stage involves the generation of the nucleocapsid (Fig 7.2, stage 3), i.e. replication of the viral genetic material and formation of the M_1 protein. The new virion assembles at the host cell surface at a point where large numbers of HA and NA molecules exist (Fig 7.2, Stage 6). A bud is subsequently formed at the cell surface (Fig 7.2 stage 7), which subsequently pinches off to release the progeny influenza virion (Fig 7.2, stage 8). This *budding* mechanism results in an influenza virion containing the same lipid bi-layer as the host cell.

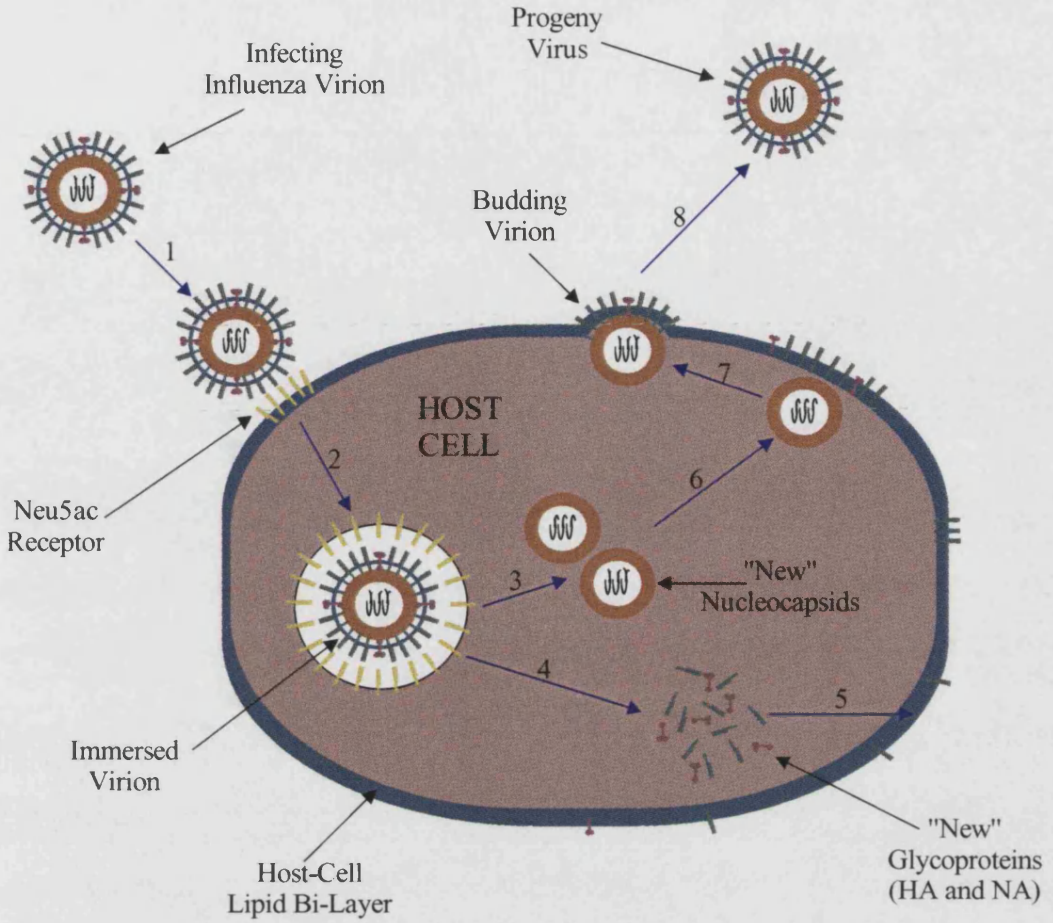


Fig 7.2. The basic stages of influenza virus replication in a host respiratory-tract epithelial-cell. An infecting influenza virion attaches to the cell in which the HA recognises terminal Neu5ac residues (yellow) on the cell surface. The virion is engulfed by the cell, resulting in a virion-receptor complex (stage 2). The virus and cell membranes fuse, allowing replication of the genetic material and eventually formation of new nucleocapsids (Stage 3). At the same time new glycoproteins are synthesised (stage 4) and are transported to the host-cell surface (Stage 5). When formed, the new nucleocapsids transfer to the cell surface at a point where the concentration of HA and NA molecules is high (Stage 6). Here, the progeny virus is assembled forming a bud (stage 7) that eventually pinches off the host cell to give the free progeny virion (Stage 8).

The function of NA is to catalyse the removal of terminal sialic acid residues resulting in the release of the Neu5ac product. This action has several possible roles in the influenza life cycle. These include removal of the terminal sialic acids from the host-cell surface during viral replication and subsequent aiding of the budding process,[109] as well as prevention of virion agglutination by removal of terminal Neu5ac residues from both surface glycoproteins.[110] The enzyme may also facilitate the initial approach of the virion to the host epithelial cell, since such cells are covered with a layer of mucin.

The continued infectious activity of influenza-A can be attributed to the continuous mutations in both antigen (HA and NA) structures.[104] Major mutations in either glycoprotein can result in structures with only 50% homology, giving a new subtype of HA or NA. Currently, 13 subtypes of HA (H1 – H13) and nine subtypes of NA (N1 – N9) are known.[104] Usually, only influenza viruses with the N1 and N2 sialidase subtypes infect humans. Mutations leading to new subtypes of antigen are usually responsible for pandemics but conservative mutations involving a few residues often lead to epidemics.[104]

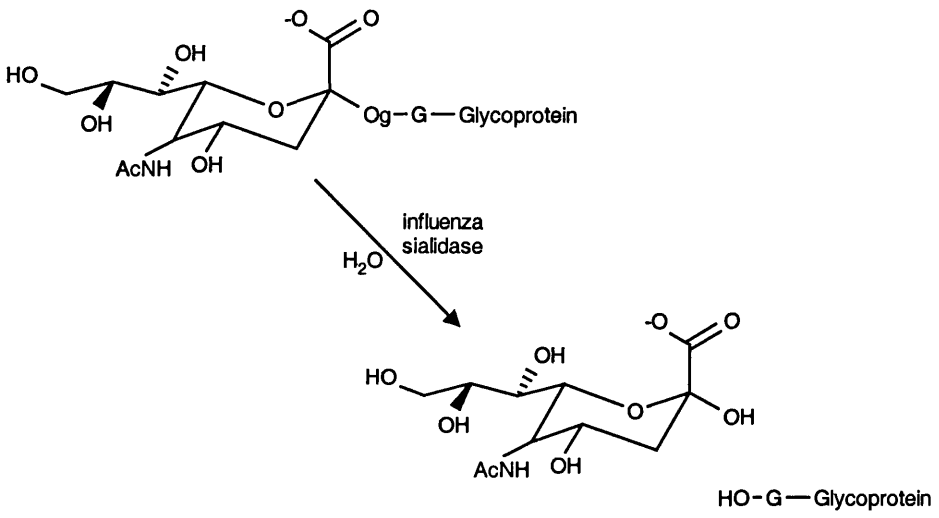
Therefore, there exists several possible ways to hinder the onset of influenza infection. Firstly, prevention of RNA replication can be achieved by using anti-virals such as amantadine and rimantidine.[104] Unfortunately, these therapeutics have unwanted side effects and are also ineffective against the influenza-B virus; in addition, resistant virus strains appear quickly. Alternatively, vaccinations against influenza are often effective; however, it is difficult to predict influenza types to vaccinate against on a year-by-year basis.

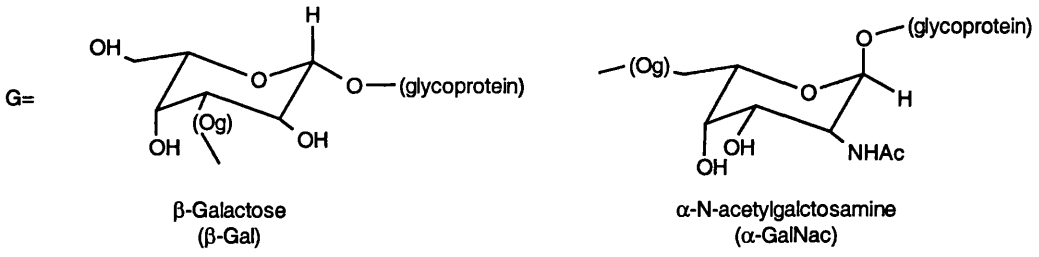
Another approach is to hinder the mechanism of either surface glycoprotein. Whitesides and co-workers have succeeded in inhibiting HA action by using a poly(acrylic acid) polymer containing Neu5ac side chains.[111,112] The sialic acids bind to the glycoprotein while the polymer mesh prevents HA's approach to the host cell. However, the inhibition of NA would hinder the motion of virus progeny through mucin as well as result in the agglutination of virions, resulting in a lower number of host cells being infected. Thus, the inhibition of sialidase action is a plausible therapeutic target and ways of inhibiting this enzyme are discussed in section 7.3, after a brief introduction into the NA's structure and function.

7.2. Structure and Enzyme Action of Sialidase (NA)

The sialidase enzyme on the influenza virion surface consists of a boxed-shaped tetrameric head (100 x 100 x 60Å) that is anchored just below the surface of the viral membrane by a "knob". Each monomer in the head has an active site, a small crater on the surface of the sub-unit that recognises the sialic acid receptor. The molecular weight of the whole enzyme is about 250kDa, with each monomer head being about

50kDa.[113] The function of sialidase is to catalyse the hydrolysis of glycosidic links involving terminal Neu5ac residues on oligosaccharides, glycolipids and glycoproteins, resulting in the release of the sialic acid product (scheme 7.1). The glycoprotein on the host epithelial cell surface contains 16 oligosaccharide sites with Neu5ac residues bonded to either galactose or N-acetylgalactosamine residues via an α-2,3 and α-2,6 glycosidic link respectively.[106,114]





Scheme 7.1

The structure of the enzyme sub-unit consists of six four-stranded anti-parallel β-sheets arranged analogously as blades of a propeller (figure 7.3).[115] Each sheet has a right-handed twist with the first and fourth strands being almost at right angles, the first strand being almost parallel with the four-fold axis of the enzyme head. The first strand of each sheet (located at the centre of the sub-unit) is connected to the fourth strand of the previous sheet by a loop that runs across the top of the sub-unit. The N-terminus, which almost runs along the four-fold symmetry axis, is found at the base of the sub-unit; it builds the fourth strand of the sixth sheet before progressing on to

the first-strand of the first sheet.[115] To date, the structures of influenza-A N2[116] and N9[117,118] sialidase have been solved to 2.2Å[116] and 2.5Å[118] resolution respectively while N5, N6 and N8 subtypes have been crystallised;[119] influenza B NA has been resolved to 2.2Å.[120] Essentially, these studies showed that all influenza NA sub-types have the same protein fold.

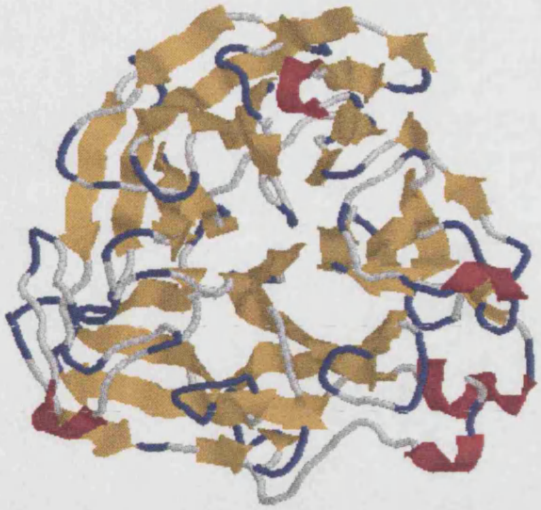


Fig 7.3. Secondary structure of a single NA sub-unit (as viewed from the top). Yellow parts of the protein represent β-sheets; red parts represent helical twists; blue parts represent β-turns.

Crystal structures have also been determined for influenza sialidase–substrate complexes. Varghese *et al.* have published structures that incorporate either the Neu5ac product or the 2-deoxy-2,3-dehydro-N-acetyl Neuraminic acid (Neu5ac2en, **2**) inhibitor.[24] In addition, Burmeister *et al.* have published NA structures with bound Neu5ac or Neu5ac2en substrates at 1.8Å resolution and have shown that Neu5ac binds into the active site in the $B_{2,5}$ boat conformation.[121] The structure of Varghese *et al.* is also consistent with this mode of binding, with the carboxylate adopting a pseudo-equatorial position.[24] The other side chains of the substrate also adopt the equatorial position in this conformation. The N9 sub-type complexed with the Neu5ac2en substrate has also been solved to 2.8Å resolution.[118]

From the influenza-B NA study, Burmeister *et al.* undertook a series of experiments in $H_2{}^{18}O$ and showed that NA was able to turnover labelled Neu5ac from the original product.[121] Under non-enzymic conditions, there was no evidence for this reaction; the labelled product obtained under enzymic conditions was suggested to have resulted from solvent attack upon a sialyl zwitterion intermediate.[121] In another experiment, soaking NA crystals in sialyl lactose for about 20 hours resulted in no electron density being obtained for the O2 oxygen.[121] Burmeister *et al.* concluded from MS/GC analysis that sialidase was able to eliminate water from the Neu5ac product to yield the Neu5ac2en inhibitor in a non-reversible side reaction.[121]

From the enzyme-substrate solved structures, some 18 active site residues were found to be strictly conserved throughout the influenza sialidase family (N1 –N9 and B), even though sequence homology between structures may only be as much as 30%. This includes all active site residues which directly interact with the substrate (figure 7.4).[24,115-118, 120,121]

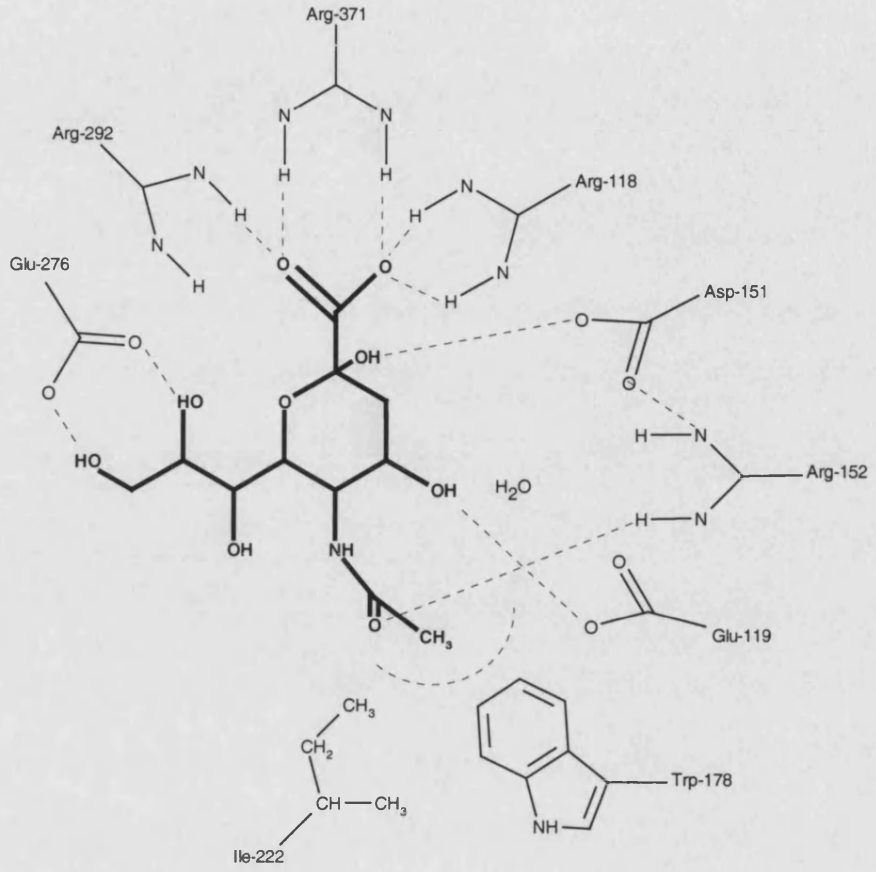


Fig 7.4. Interactions between conserved residues in influenza sialidase and Neu5ac substrate (bold).

The substrate carboxylate group, located at the 2-position, interacts with three arginine residues of the enzyme (Arg-118, Arg-292 and Arg-371). Asp-151 lies on top of the active site making a hydrogen bond to the 2-hydroxyl group. Asp-151 also interacts with Arg-152, the latter making a hydrogen bond to the carbonyl oxygen in the substrate N-acetyl side-chain. Glu-119 interacts with the 4-OH group, with the resulting "empty pocket" being filled with a water molecule. The methyl group within the substrate N-acetyl side-chain makes hydrophobic, non-bonded interactions with the Ile-222 and Trp-178 residues of the enzyme. The "outer" two hydroxyl groups on the glyceryl side-chain interact with the Glu-276 residue.

On the floor of the active site, Glu-277 makes a hydrogen bond to Arg-292, while the other oxygen interacts with the hydroxyl group of Tyr-406. The Tyrosine hydroxyl oxygen is located about 3.0Å directly below the C2 carbon of the substrate. These two residues are also conserved in all solved NA structures.

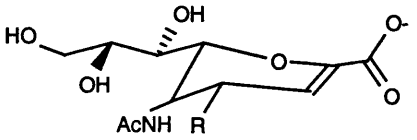
Thus, from structural studies, influenza NAs have been shown to possess an active site that is strictly conserved in more than one sub-type.[24,115-118,120,121] A sialidase inhibitor that hinders the catalysed reaction by binding to the active site would therefore be an effective therapeutic against all strains of influenza. Considering that any structural changes in either glycoprotein often leads to epidemics and pandemics, great interest in designing effective NA inhibitors has ensued.

7.3. Inhibitors of Influenza Sialidase

Knowledge of the 3D structure of influenza NA has been used in conjunction with molecular modelling to generate several classes of potent inhibitors for the enzyme. von Itzstein *et al.* have demonstrated the applicability of this approach by surveying the active site of a determined sialidase structure for potential interactions between the substrate and the protein.[122] Such simulations roll different "chemical probes" around the active site and elucidate interactions between the probing substituent and the enzyme.[122] Thus, for example, a carboxylate probe interacts well with the arginine triad shown in figure 7.4.

This process of *rational drug design* has provided knowledge of a negatively charged pocket about the substrate 4-position, which could conceivably interact with a positively charged group at this substrate site. Substituting the OH moiety in Neu5ac2en for ammonium or guanidinium groups resulted in substrates with increased potency to hinder enzyme action.[123] Both 4-amino-2,4-deoxy-2,3-dedihydro-N-acetyl neuraminic acid (**3**, Am4Neu5ac2en) and 4-guanidino-2,4-deoxy-2,3-dedihydro-N-acetyl neuraminic acid, (**4**, Gu4Neu5ac2en) have inhibition constants (K_i) in the nanomolar range and are strong inhibitors of NA.[123] In comparison, Neu5ac2en, **2**, has a K_i in the micromolar range and is a moderate inhibitor *in vitro*. Crystal structure analysis of the sialidase-Am4Neu5ac2en complex

gave evidence for the formation of a salt-bridge between Glu-119 and the ammonium group.[123] Replacing the 4-ammonium substituent with a bulky 4-guanidinium moiety results in increased favourable interactions with the pocket around the 4-position: both NH_2 sites interact with either Glu-119 or Glu-227.[123] In addition, Asp-151 and the backbone carbonyl oxygen of Trp-178, which complete the definition of the pocket, also form favourable interactions with the 4-guanidinium group. When Neu5ac2en is bound to NA, a water molecule occupies the binding pocket surrounding the 4-OH group. Therefore, an entropic contribution may also assist the binding of Gu4Neu5ac2en to sialidase: the "pocket water" molecule must be displaced to allow binding of the inhibitor.[123]



2 R= —OH

3 R= —NH_3^+

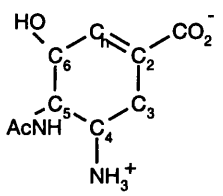
4 R= —NH-C(=NH_2^+)-NH_2

Structure activity relationships (SAR) on Gu4Neu5ac2en have shown that addition of further functionality to the 4-guanidinium group lowers the activity of the inhibitor, indicating a subsequent increase in steric hindrance at the 4-position.[124] The N-acetyl group activity has also been investigated, but larger groups again introduce steric hindrance while smaller groups eliminate the hydrophobic contacts made to Trp-178 and Ile-222.[125] Removing the N-acetyl group altogether also results in the reduction of inhibitory activity.[126] The glyceryl side chain at the 6-position is also important for the activity of Gu4Neu5ac2en. Removing the chain completely results in significant activity loss that is gradually regained as the chain is restored (one carbon centre at a time) to its original length.[127] This emphasises the importance of the interaction between substrate and the Glu-276 residue. Replacement of pyranose oxygen with a methylene group has little effect on the activity of Gu4Neu5ac2en analogues.[128]

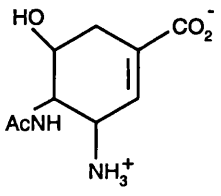
These SAR studies therefore suggest that Gu4Neu5ac2en achieves good inhibitory activity in two ways: firstly, interactions between the inhibitor and enzyme are increased through the 4-guanidinium group. Secondly, essential interactions found between the enzyme and Neu5ac substrate are retained by the inhibitor.

Intranasal administration of **4** into post-infected ferrets led to a full recovery from several influenza-A strains and influenza-B.[123] Subsequent phase-II and phase-III clinical trials have been successful in humans.[129] The therapeutic, with a generic name, zanamivir, is currently under regulatory review in the US and the EU while reviews have already been successful in Australia and Sweden; it is likely to be marketed under the name of Relenza™.[130] Because the therapeutic has low bioavailability when administered orally, it must be delivered to the respiratory tract via intranasal administration.[123]

Kim *et al.* subsequently designed an orally active sialidase inhibitor by reducing the hydrophilicity of the molecule as compared with zanamivir.[131] They considered a chemically versatile cyclohexene scaffold on which to add pendent functionalities. The cyclohexene ring contained a double bond between the C2 and C_h atoms, the same place in which conjugation in the sialyl zwitterion takes place. The importance of having the double bond in this position is emphasised by recognising the generally accepted rule that an enzyme catalyses a reaction through stabilisation of the TS structure.[147] An effective inhibitor would therefore resemble the TS structure for the reaction.[148] Many groups have suggested that the TS for sialidase catalysed sialoside hydrolysis takes place via a "sialyl zwitterion like" TS (Section 8.2). Using a model system in which the glyceryl side chain was replaced with a hydroxyl group, Kim *et al.* showed that **5** gave a structure with greater inhibition activity than **6**.[131]

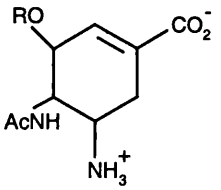


5



6

After deciding to use the carbocyclic scaffold of **5**, Kim *et al.* subsequently replaced the glyceryl side-chain with an ether function.[131] Their reasoning for this was two fold. Firstly, replacement of the glyceryl side chain results in a more hydrophobic molecule and secondly, the ether link replaces a hydroxyl group that doesn't interact with the sialidase enzyme. They also suggested that electron density of the double bond might be reduced through electron withdrawing effects (via the σ framework) associated with the ether oxygen.[131] The carboxylate and acetamino groups are retained at the 2- and 5-positions respectively, with an ammonium group placed at the 4-position to interact with the adjacent negatively charged pocket. Kim *et al.* measured the potency of a series of inhibitors based around this framework.[131] They discovered that on lengthening the "ether chain" (from R=methyl, **7**, to R=*n*-propyl, **9**), inhibitory activity of the molecule was increased, while R=*n*-butyl, **10** decreased the inhibitor's ability to inactivate the enzyme. Thus, R has limited space in which to pack within the active site. Branched alkanes showed even greater binding affinity, especially when R=3-pentyl, **11**. This resulted in *ca.* 200-fold increase in the inhibitory activity over **9**. X-ray crystal structures suggested that these hydrophobic chains are accommodated in the active site by forcing Glu-276 (which normally interacts with the glyceryl side chain) to "orientate outward" with subsequent formation of a hydrophobic pocket.[131]

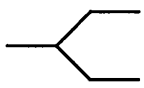


7 R= —CH_3

8 R= —CH_2CH_3

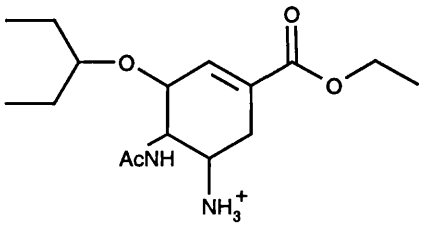
9 R= —$CH_2CH_2CH_3$

10 R= —$CH_2CH_2CH_2CH_3$

11 R= 

SAR studies on **9** showed that replacing the ether oxygen with sulfur or methylene function results in little reduction in the molecule's inhibitory activity, suggesting the changes in electronic effects at this site are not important for sialidase inhibition.[132] Extending the hydrophobic chain on the N-acetyl group also results in a reduction in enzyme inhibition.[133] Replacing the 4-ammonium substituent with a guanidinium group gives a molecule with *ca.*70-fold increase in activity, probably accompanied by reduced oral bioavailabilty.[133] Subsequent addition of a methyl group to the 4-guanidinium substituent results in a slight decline in activity.[133] It has also been noted that while **9** benefits from a 70-fold increase in activity when the 4-ammonium moiety is replaced with a guanidinium group, **11** shows only a 2-fold increase.[133] This suggests that improvements on enzyme inhibition obtained with individual structural changes are not directly co-operative. Finally, in another study, adding various substituents to C_h was found to markedly reduce its inhibitory activity of **11**.[134]

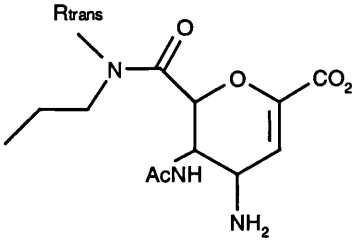
From these studies, carbocyclic analogue, **11**, was put forward as a possible therapeutic inhibitor of sialidase and was designated GS 4071; this had poor bioavailability and was subsequently modified to the ethyl ester (designated GS 4104, **12**). The inhibitory activity of GS 4071 on influenza-A sialidase is comparable to that of zanamivir.[135]



12

Smith *et al.* have subsequently replaced the glyceryl side chain on zanamivir with an amide function possessing two N-bonded alkyl groups with the view of designing an orally active influenza therapeutic.[136,137] In the same manner as GS 4104, the binding of *carboxamide* inhibitors to NA results in the formation of a salt bridge between Glu-276 and Arg-224. The ensuing hydrophobic cavity is large enough to contain an

n-propyl group (located *cis* to the pyranose ring in the carboxamide), while larger aliphatic groups result in steric hindrance and some loss of inhibition activity. X-ray crystal structures have shown that the *trans* group can be accommodated by a long, thin hydrophobic cleft, defined by Ala-246 and Ile-222 residues.[137] Extensive modification of the R_{trans} group has little effect on the inhibitor's activity: comparable inhibition to zanamivir is obtained, for example, when R_{trans} is an *n*-propyl, **13**, or phenylethyl, **14** substituent.[136-138]



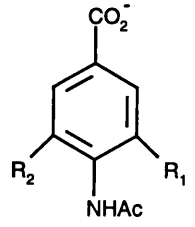
13 R_{trans} = $-CH_2CH_2CH_3$

14 R_{trans} = $-CH_2-CH_2-C_6H_5$

Taylor *et al.* have used structural and modelling studies to account for the apparent selectivity of carboxamide inhibitors towards influenza-A NA over the B-sialidase.[139] From X-ray studies, it was shown that a slight expansion and distortion in the influenza-B enzyme structure was required for the formation of an Arg-224 – Glu-276 salt bridge. While Glu-276 makes a bi-dentate interaction with the sialic acid substrate in influenza-A and influenza-B sialidases, the conformation of its side chain differs between the two enzymes. In binding the carboxamide to the B-sialidase, an energetic penalty may have to be paid in order to place the Glu-276 in the correct formation for salt-bridge formation. Interestingly, MD simulations have shown that while the majority of low energy structures for the inhibitor-sialidase(A) complexes contain the Glu-276 - Arg-224 salt-bridge, a lower proportion of the influenza-B sialidase complexes contained this interaction.[139] Furthermore, R_{trans} was more frequently observed to "pop out" of its hydrophobic cleft than in the sialidase-A enzyme.

Finally, substituted benzoic acids have also shown some mild inhibitory activity towards influenza sialidases.[140-143] The driving force behind aromatic inhibitors is the ease of synthesis compared to the stereo-demanding sialic acid template found in the zanamivir molecule. Benzoic acid templates that contain any group at the notional

5-position other than acetamino substituent showed little inhibitory activity.[143] The crystal structure of bound *p*-acetaminobenzoic acid, **15**, to sialidase showed that the acetamino and carboxylate species adopt a similar position in the active site compared to Neu5ac2en.[143] Placing a guanidinium group at the meta-position, **16**, in order to mimic the 4-guanidinium group in zanamivir, resulted in the basic substituent binding into a position usually occupied by the glyceryl side chain.[141,144] It has been suggested that (in as far as benzoic acids are concerned) this orientation may present the guanidinium group in a more favourable binding site.[141,143] Sudbeck *et al.* noted that this binding mode results in another carboxylate arginine-triad motif (involving the guanidinium group on the benzoic acid, Arg-224, Arg-292 and the Glu-276 carboxylate group).[144] They suggested that this extra arginine-carboxylate motif might be responsible for the inhibitory activity of **16**. The addition of a further guanidinium group to the notional 4-position, **17** reduces the inhibitory activity of the substituted benzoic acid.[144] The activity of **16**, the most potent benzoic acid derivative inhibitor of sialidase, is comparable to Neu5ac2en.



	R_1	R_2
15	—H	—H
16	—NH—C(=NH$_2^+$)—NH$_2$	—H
17	—NH—C(=NH$_2^+$)—NH$_2$	—NH—C(=NH$_2^+$)—NH$_2$

In conclusion, active research is currently under progress to find novel sialidase inhibitors that could be administered as effective therapeutics against influenza. This includes Neu5ac2en analogues (including zanamivir), carbocyclic analogues (such as GS 4104), carboxamides and substituted benzoic acid inhibitors. It is essential that progress continues in this field since sialidase mutants are likely to arise when the

enzyme is confronted with such inhibitors. Mutants have indeed been successfully grown in the presence of zanamivir *in vitro*.[145,146] An amino acid change from Glu-119 to Gly results in reduced activity of the inhibitor by reducing the favourable interactions associated with the 4-guanidinium group. Glu-119 is not essential for Neu5ac binding and this mutation thus provides a facile way to overcome inhibition. It is well accepted that enzymes catalyse reactions by lowering the energy of the TS structure for a reaction.[147,148] If TS-analogue inhibitors could be designed for sialidase, then any subsequent mutation would probably effect the enzyme's efficiency. However, to be able to design a TS analogue requires knowledge of the experimentally elusive TS structure.[148] The exact sialidase mechanism is not known and molecular modelling is now in a position to simulate such reactions, leading to subsequent insight on the TS structure associated with the rate-determining step.

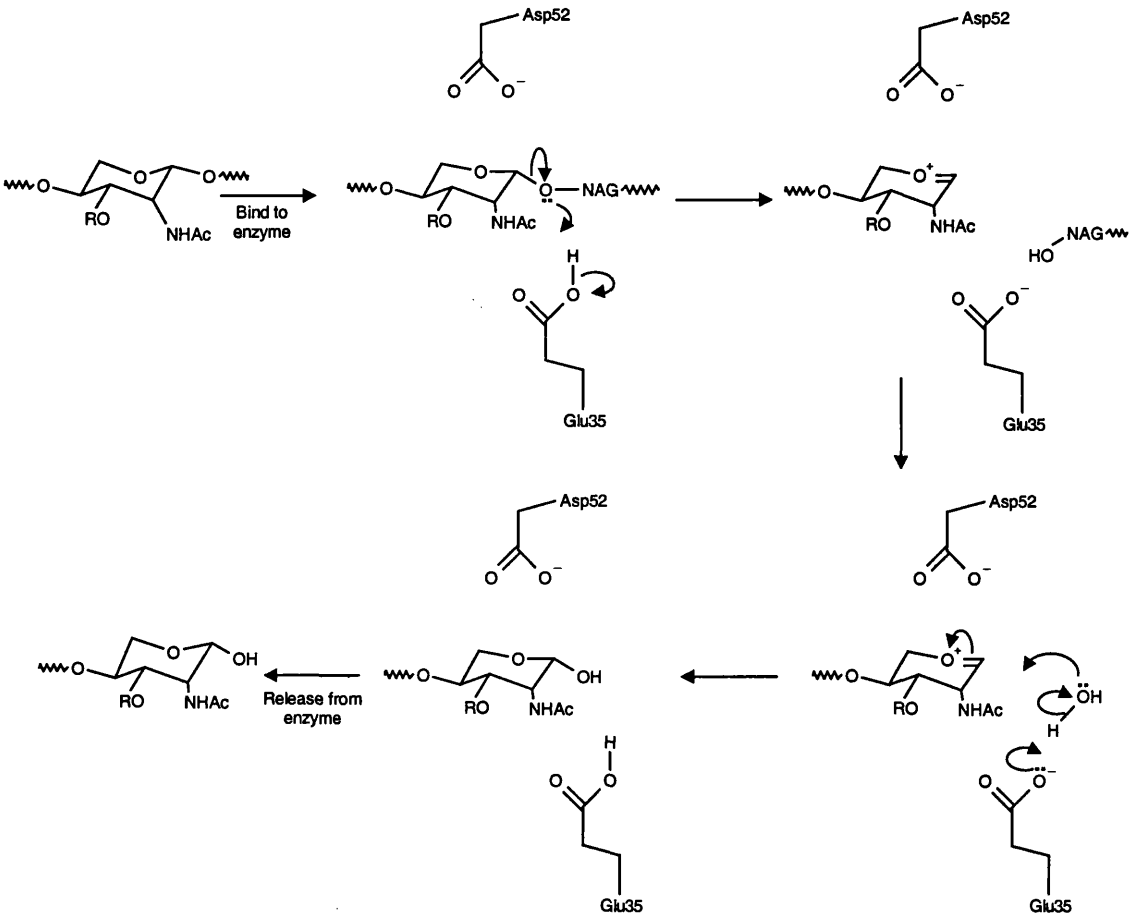
8. QM/MM HYBRID MODELLING OF INFLUENZA-A SIALIDASE ACTION ON SIALOSIDES

8.1. Introduction

8.1.1. Insights into Enzyme Mechanisms Using Hybrid QM/MM Techniques

QM/MM techniques have made investigations into enzyme catalysed reactions computationally amenable and have provided further mechanistic insight into such processes. Warshel and Levitt proposed a hybrid methodology that allowed them to investigate mechanistic aspects of hen egg-white lysozyme (HEWL) catalysed hydrolysis of glycosides.[149] For this study, they combined a semi-empirical Hamiltonian description of the reaction centre with a united atom force field description of the remaining enzyme atoms. Solvent effects upon this system were described by assigning dipoles to lattice points, rather than treating solvent molecules as discrete entities. Throughout the simulations, the MM part of the enzyme remained largely fixed while allowing geometry optimisation of the QM system. The 15-atom strong QM region included the majority of the substrate pyranose ring and part of the protonated Glu-35 active site residue.

The Phillips mechanism (scheme 8.1) of the HEWL catalysed hydrolysis of glycosides[150] first involves substrate binding to the enzyme active site in which the glycoside ring distorts towards a half-chair conformation. This weakens the glycosidic link that is subsequently cleaved through general acid catalysis involving the adjacent Glu-35 residue. The TS structure for this process has been suggested to have positive charge development on the glycoside ring that is stabilised by an aspartate residue (Asp-52) located on the opposite side of the active site to Glu-35. After cleavage of the glycoside bond, a water molecule attacks the oxocarbenium ion intermediate and the resulting product then leaves the active site.

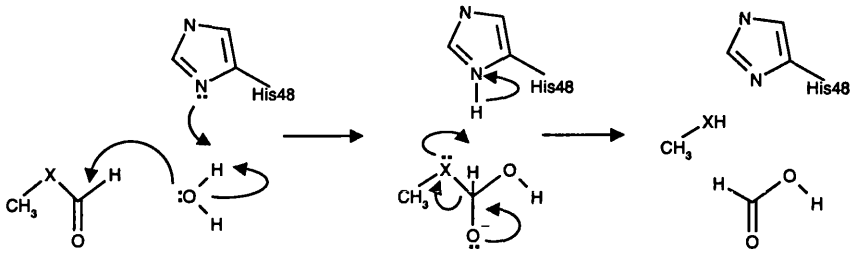


Scheme 8.1

Warshel and Levitt's study found that Asp-52 stabilised the oxocarbenium ion by around 38 kJ mol^{-1} over the bound reactant-state. Furthermore, the enzyme afforded the oxocarbenium ion some 165 kJ mol^{-1} stabilisation compared to the same species in the gas phase.[149] These results showed how effective the molecular mechanics potential was in providing some electrostatic stabilisation to the QM system. From this study, it was also concluded that electrostatic strain was of greater importance than steric strain in the HEWL catalysed mechanism.

Later, Waszkowycz *et al.* investigated the phospholipase-A2 (PLA2) mechanism on esters[151] (scheme 8.2, X=O) and amides[152] (scheme 8.2, X=NH). The role of PLA2 is to catalyse the hydrolysis of the 2-acyl-ester bond in L-α-phospholipids found on cell membranes; this process results in the release of fatty acids. The QM/MM simulation set-up was somewhat different to Warshel and Levitt's work. The model

contained some 200 atoms that included model substrate, 17 amino acids from the protein, a calcium ion and four water molecules.



1

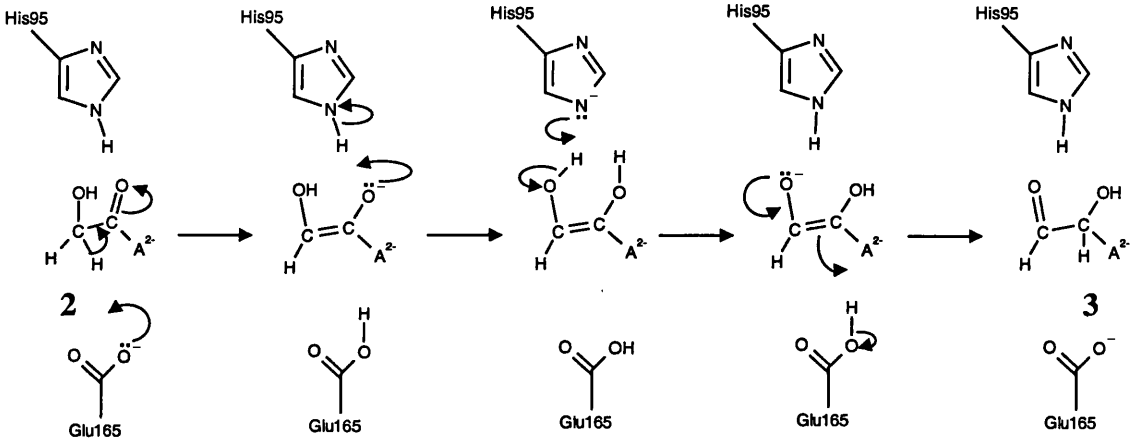
Scheme 8.2

The QM region for this system consisted of methyl formate (a model for the ester part of L-α-phospholipids), a water molecule and the side chain of His-48. These 20 QM atoms were described at the HF/3-21G level of theory within a fixed field of MM point charges and van der Waals spheres.[151] Waszkowycz *et al.* used an adiabatic mapping procedure: from the oxyanion intermediate, **1**, the C—OH and C—OMe bonds were elongated in 0.1Å steps towards reactants and products respectively. In this study, Waszkowycz *et al.* found the enzyme almost removes the barrier for proton transfer between the incoming water molecule and His-48 compared to the same change *in vacuo*.[151]

The same authors also investigated the PLA2 catalysed hydrolysis of amides, carried out in essentially in the same way as the ester hydrolysis modelling.[152] In this case however, Waszkowycz *et al.* were able to characterise a TS structure for the amide hydrolysis, the barrier resulting from the destruction of planarity of the amide link.

Although these simulations by Waszkowycz *et al.* allowed the combination of the MM environment with an *ab initio* description of the reacting centre,[151,152] they suffer from several drawbacks. Firstly, the size of the MM environment surrounding the QM reacting centre is small: longer-range electrostatic interactions of the protein may have been neglected. Secondly, the flexibility of the system was very limited; however, the use of a smaller, manageable model necessitated the use of a fixed MM region.

An alternative approach is to combine a less computationally intensive description of the QM system (by using a semi-empirical Hamiltonian) with a larger number of MM atoms surrounding the QM centre. Bash *et al.* investigated the interconversion of dihydroxyacetone phosphate (DHAP, **2**) to D-glyceraldehyde 3-phosphate, **3**, catalysed by the triosephosphate isomerase (TIM) enzyme[153] (scheme 8.3, A=CH_2OPO_3). The postulated mechanism first involves proton abstraction at the 1-position by Glu-165 and proton donation to the resulting oxyanion by His-95. The same His residue then abstracts a proton from the substrate 1-hydroxyl group, which is followed by proton transfer from Glu-165 to the former carbonyl carbon that results in the enzyme-bound product, **3**.



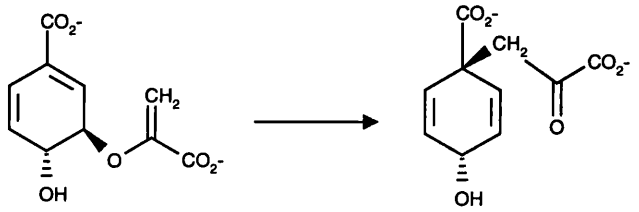
Scheme 8.3

For this study, all protein residues containing atoms within a 16Å radius of the active site were included in the model, along with 100 water molecules added to solvate the ES complex. The His-95 and Glu-165 side chains were included in the QM region, along with the substrate, resulting in a model with 35 QM and 1300 MM atoms. The proton transfers were mapped out in 0.2Å stages from a starting structure that had been energy minimised to an RMS gradient tolerance of about 0.04 kJ mol^{-1} $Å^{-1}$. Each step on the profile was minimised for 250 steps.[153]

Bash *et al.* found that unfavourable DHAP-Glu165 interactions in the active site destabilised the bound reactant compared to the unbound DHAP.[153] Furthermore, the enzyme was found to stabilise the reactant anion and the enediolate intermediate (with respect to the same structures in the gas phase). The enediolate was stabilised

by a greater amount, reducing the reactant-intermediate energy difference compared to that calculated in *vacuo*. In a further study, they also noted that in the course of the proton transfer between protonated His-95 and enediolate anion, there was a large drop in the energy of the system (about 75 kJ mol^{-1}). This resulted from a pK_a imbalance between the protonated His-95 and the enediolate anion, affording greater stability for the bound intermediate anion. It was suggested that such stable intermediates are likely to hinder the reaction rate.[153]

In another study, the enzyme catalysed conversion of chorismate, **4** to prephenate, **5** via a pericyclic process was considered[154] (scheme 8.4). It is thought chorismate mutase (CM) doesn't take part in the reaction mechanism directly but enforces a reacting conformation of the chorismate reactant (in this case the diaxial conformer). Lyne *et al.* have studied this process using hybrid techniques in which 1800 protein atoms and 85 water molecules were modelled. This used a similar methodology to that of Bash *et al.* by only including protein residues within 16Å of the active site. The outer 4Å region of the protein was subjected to a harmonic force constraint. Several QM partitions were chosen: one case only included the substrate in the QM region but other models included active site residues of CM. However, all models gave essentially the same results.[154]

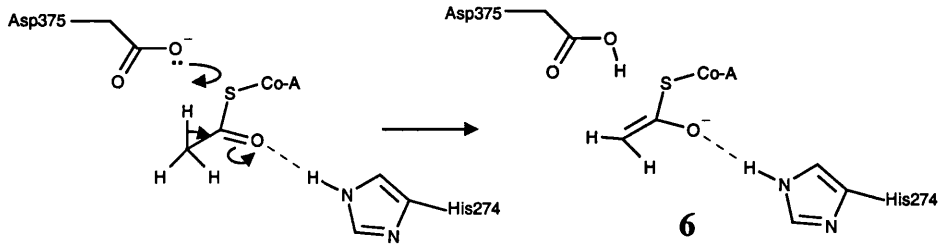


4 **5**

Scheme 8.4

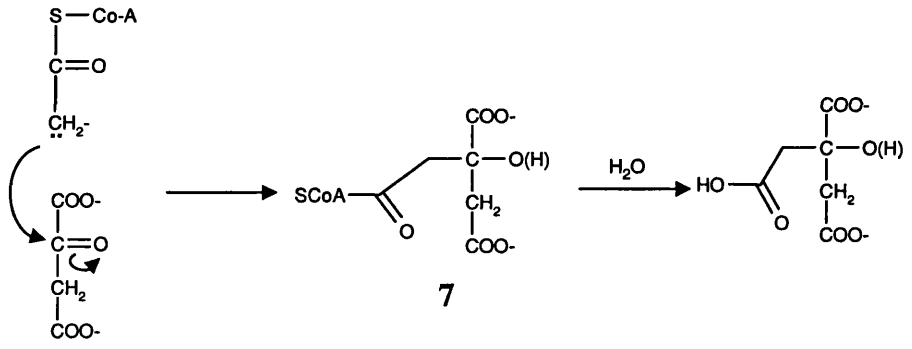
Lyne *et al.* discovered the barrier for the pericyclic process was reduced by approximately 100 kJ mol^{-1} for the catalysed reaction compared to the same process *in vacuo*. It was found that CM achieves this reduction by stabilising the TS structure through hydrogen bonds between active site residues (Arg90 and Glu78) and the substrate. They also suggested that binding a strained conformation of the substrate to the enzyme also contributes to the reduction in the barrier height.[154]

Mulholland and Richards used a hybrid QM/MM methodology to investigate the rate-determining step (RDS) for the formation of citrate from acetyl co-enzyme-A (acetyl-CoA) and oxoloacetate.[155] The RDS of the mechanism (catalysed by citrate synthase) is thought to involve proton abstraction (by an active site base) from the acetyl Co-A to give the enolate anion, **6** (Scheme 8.5).



Scheme 8.5

This is followed by addition of the resulting enolate anion to the carbonyl centre of the oxaloacetate (Claisen condensation), resulting in the formation of citryl-CoA, **7**. Finally, the addition of water gives the citrate product (Scheme 8.6).

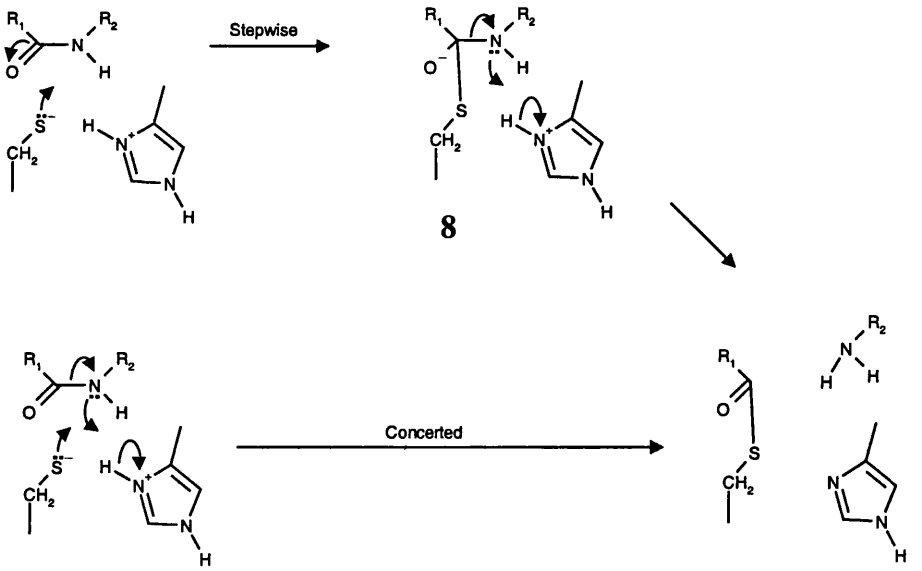


Scheme 8.6

The QM/MM system contained 2600 atoms; it included any enzyme residue found within a 20Å radius of the active site. Some 33 QM atoms were treated using the AM1 semi-empirical Hamiltonian: this included the thio-ester portion of the Co-A and the side chains of Asp-375 and His-274. The MM region was described using a united atom force field. Mulholland and Richards mapped the proton transfer between the acetyl Co-A and Asp-375 in 0.2Å steps to yield the enolate

intermediate.[155] This study found that Asp-375 could act a base, with the resulting enolate being stabilised by His-274.

Hillier's group has subsequently used a semi-empirical QM/MM approach to investigate an amide hydrolysis reaction catalysed by the enzyme, papain.[156,157] The mechanism is thought to involve as Cys^{-}-His^{+} ion pair in the active site but it is not clear whether the mechanism is concerted or step-wise, the latter involving a tetrahedral intermediate, **8** (scheme 8.7). As with the earlier phospholipase-A_2 work, the original substrate was replaced with a much simpler system; this restricts the simulation of a flexible MM environment. The work involved treating 53 atoms with the semi-empirical AM1 Hamiltonian (this included model substrate, Cys-25 and His-159 side chains) within a fixed MM environment of 3500 atoms. Amber[158] and G94[8] were coupled-together for this investigation but the implementation prohibited simultaneous optimisation of both QM and MM regions. Twelve of the QM atoms were fixed to speed up the convergence of the optimisation.

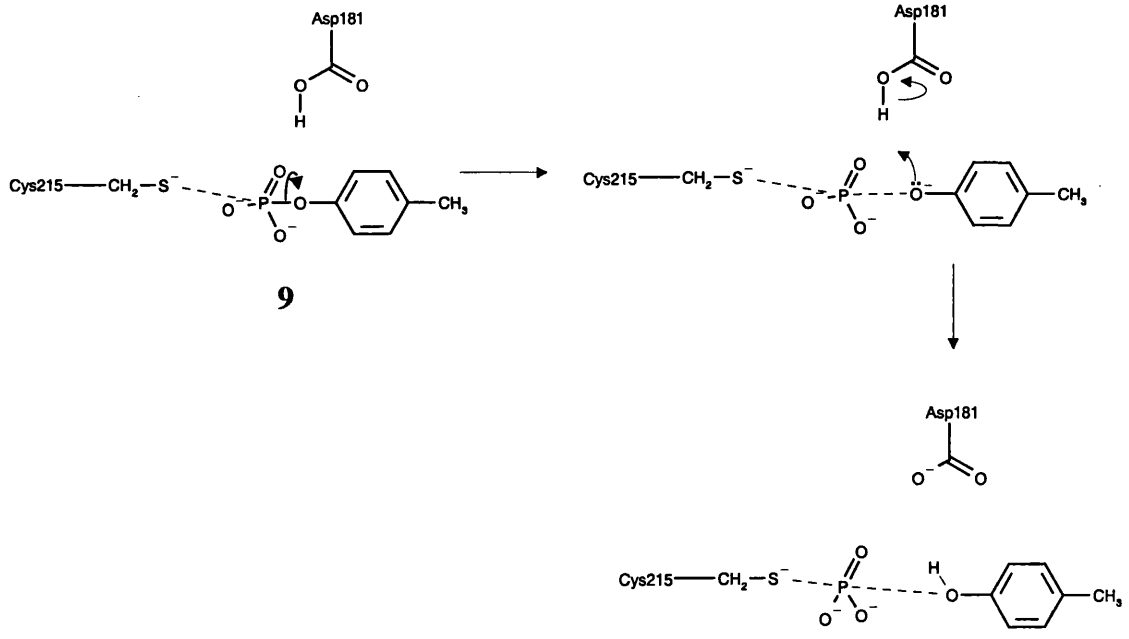


Scheme 8.7

A 2D-energy surface describing the changes in the (Cys)S—CO and (His)NH—N distances was mapped (involving the generation of 100 structures). A finer scan was done around the approximate TS structure to locate the transition-state. This structure was subsequently refined using a TS searching algorithm and confirmed as such by calculating the second derivatives of the energy.[157] However, no IRC

calculation was used to verify the nature of the TS. Within the limitations of the model, Harrison *et al.* proposed a TS structure that lies on a concerted pathway for the hydrolysis reaction.[156,157]

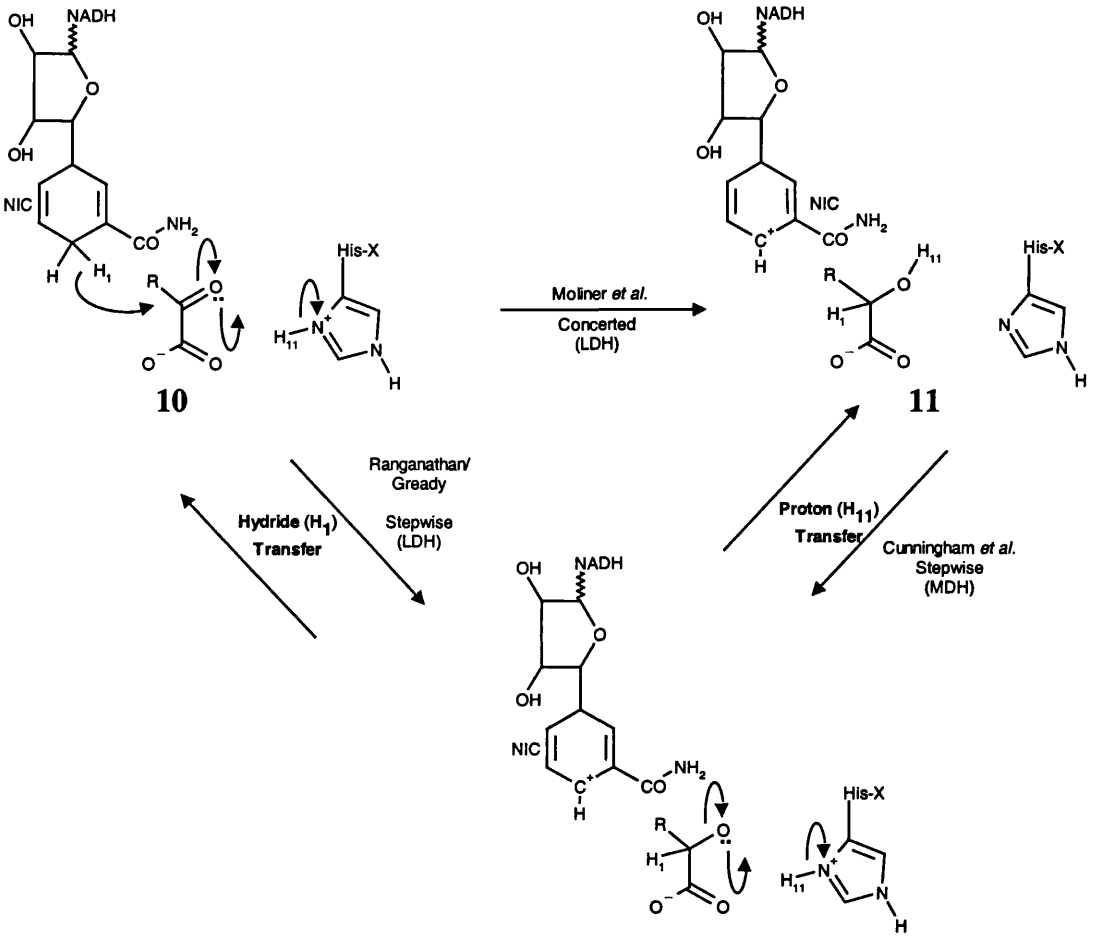
Further to these studies, Hart *et al.* also characterised a TS structure for the hydrolysis of tyrosine phosphate, **9** (catalysed by protein tyrosine phosphatase, Scheme 8.8).[26] It was suggested that the TS structure represented advanced P-O cleavage with little proton transfer from Asp-181 and little increase in the (Cys215)S—P distance. Further elongation of the P—O distance resulted in a smaller energy barrier representing the proton transfer from Asp-181 to substrate. Again, no IRC calculations were performed to confirm the nature of the TS structure for the suggested rate-determining step. Furthermore, the MM environment was also kept fixed throughout the P—O elongation process.



Scheme 8.8

Moliner *et al.* used QM/MM techniques to investigate the lactate dehydrogenase (LDH) catalysed interconversion of pyruvate, **10** to lactate, **11** (Scheme 8.9, R=Me, X=195).[159] The reaction is thought to proceed via two processes: firstly, proton transfer (PT) between the protonated His-195 residue and substrate carbonyl oxygen. Secondly, hydride transfer (HT) between the nicotinamide ring of the cofactor and

the protonated carbon on the substrate. However, it is uncertain whether the enzyme catalysed mechanism is concerted or stepwise (and if stepwise, in which order the transfer steps occur).



Scheme 8.9

This study allowed full optimisation of some 1900 atoms of which 39 were treated with the AM1 Hamiltonian. As in studies discussed thus far,[153-155] a number of enzyme residues within a certain distance of the active site were included in the simulation and the resulting outer-most protein atoms were subjected to a harmonic force constraint. The QM region comprised of the whole substrate, the side chain of His-195 and the nicotinamide ring on the cofactor. A 2D energy surface was constructed involving both proton and hydride transfers.

The TS was refined and characterised using GRACE (section 3.3) until the RMS gradient of TS structure was below 0.42 kJ mol^{-1}. However, unlike the Hillier team,

Moliner *et al.* used an IRC calculation to subsequently verify the nature of the TS structure by following the reaction path towards the local reactant and product wells.[159] The TS was found to be concerted in terms of the hydride- and proton transfers, with the proton transfer being at a more advanced stage than the HT. The advantage of this work over the characterised TS work of Hillier's team is that TS refinement takes place within a flexible environment.

Moliner *et al.* also considered another issue: the effect of the environment conformation upon the TS structure.[159] To do this, six structures were taken from a 30ps MD trajectory (at 5ps intervals) and an energy surface describing HT/PT was mapped out from each starting structure. For each of the six simulations the TS structure was refined, characterised and verified. Moliner *et al.* found that essential geometrical elements for the concerted HT/PT transition-state structure were invariant with respect to the environment.[159]

Ranganathan and Gready have also looked at the LDH catalysed conversion of pyruvate to lactate using a hybrid QM/MM potential.[160] Again, the His-195 side-chain and cofactor nicotinamide ring were included in the QM region. However, all 329 residues of the protein sub-unit were included in the calculation, with only *ca.* 150 atoms unconstrained (this includes the QM region and some MM atoms adjacent to it). Crystallographic waters were also included in the model but it was not stated whether any further solvent molecules were added to solvate the active site. The Ranganathan and Gready model has the advantage over the system used by Moliner *et al.* in that it fully accounts for all long-range electrostatic interactions in the enzyme sub-unit. However, the Ranganathan and Gready model is more hindered by very limited flexibility of the system. Furthermore, the reaction is modelled with only one set of "fixed environment" co-ordinates. Within the constraints of this system, the results from the Ranganathan and Gready study found that hydride transfer occurred before proton transfer in essentially a stepwise fashion.[160]

The contradictory results with the LDH modelling are somewhat concerning. It suggests that results are either dependent on how the system is treated, or by the nature of the QM/MM potential used to describe the system. In this case, Moliner *et*

al. used[159] the AM1/CHARMM QM/MM potential,[14] Ranganathan and Gready used[160] the AM1/AMBER QM/MM potential.[158]

Cunningham *et al.* have used a hybrid QM/MM method to look at malate dehydrogenase (MDH) catalysed conversion of malate to oxaloacetate, which is analogous to the reverse process of the LDH reaction.[161] The mechanism involves a proton transfer, in this case from the substrate to an active site histidine (Scheme 8.9, R=CH_2COO^-, X=177) and a hydride transfer from substrate to cofactor (NAD^+). For this study, 54 atoms were treated with the AM1 Hamiltonian (including the side chain of His-177 and the co-factor's nicotinamide and ribose rings). This work also reparameterised the AM1 Hamiltonian to accurately describe experimental proton affinities and *ab initio* geometries.

All enzyme residues within 18Å of the active site were included in the model. Extra solvent molecules were added to the system (already containing waters explicitly accounted for in the crystal structure) via a two-stage process. Firstly, a 20Å water droplet was superimposed on the enzyme system, with subsequent deletion of waters within 3.1Å of any heavy atom found in the enzyme-substrate complex. This process was repeated on a system which had been subjected to a period of MD (40ps), with all but the water molecules kept in a fixed position. The second stage results in a further 30 solvent molecules being added to the system. The environment of the QM system is finally represented by 3000 MM atoms.[161] During these simulations, a stochastic boundary was applied to water molecules and harmonic force constraints were applied to protein atoms beyond 16Å of the active site.

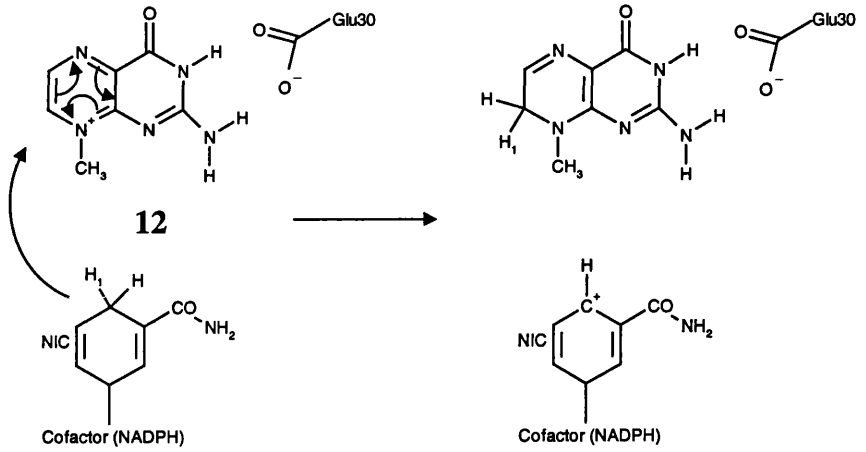
To generate the initial reference structure, Cunningham *et al.* heated the system (which at this stage was treated with MM potentials only) from absolute zero to 300K in 20ps.[161] The system was equilibrated for 80ps and data collection was subsequently taken over 40ps. From the 80ps equilibration stage, another MD phase in which the system was treated with the QM/MM potential was carried out. This part of the simulation involved 20ps of equilibration followed by 20ps of data collection. The system was then cooled from 300K to 0K over 20ps and subsequently optimised for 5000 steps. All 675 structures used to describe the

hydride- and proton transfer energy surface were generated in this manner, with PT and HT processes being mapped in 0.2Å steps.

From these simulations, it was found that proton transfer occurs before the hydride transfer, the latter being the rate-determining step of a possible stepwise mechanism. The energy barriers for these steps were calculated as 29 kJ mol^{-1} and 63 kJ mol^{-1} for the PT and HT processes respectively.

The study by Cunningham *et al.*[161] is probably the most advanced work undertaken thus far in modelling enzyme mechanisms using hybrid QM/MM techniques. It samples conformation space at 300K using MD with both MM and QM/MM potentials, it also optimises structures using a large number of optimisation cycles (*ca.* 5000). The PT and HT processes are adequately sampled by calculating a large number of structures associated with both stages. However, it should be noted that the resulting potential energy surface was obtained for a system at 0K, and barriers are only approximate because the TS structures have not been refined, characterised or verified.

Another problem in modelling reaction mechanisms is the difficulty in calculating the free energy surface for a reaction. The calculation of free energy is computationally expensive, given that conformation space must be adequately sampled. However, Cummins and Gready have recently published work on the QM/MM modelling of dihydrofolate reductase catalysed reduction of 8-methylpterin, **12** in which a hydride ion is transferred from the cofactor to the substrate (Scheme 8.10).[162]

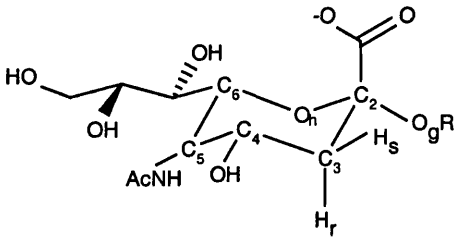


Scheme 8.10

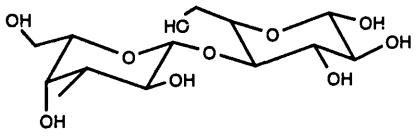
This work used a 16Å sphere for the description of the enzyme and a 22Å sphere of solvent water molecules. Four different QM/MM models were investigated, which differed in the number of cofactor residues included in the QM region. A QM/MM potential was applied to free energy calculations in which some 160 windows were calculated. The energy change associated with each window was evaluated by using 800fs of equilibration followed by 800fs of data collection. The total time-period for the simulation was thus 256ps. The free energy barrier calculated for the hydride transfer was about 125 kJ mol^{-1} and the total free energy change for the reaction was calculated as 41 kJ mol^{-1}. The positive value obtained for the reaction free energy was largely due to the favourable interaction of the protonated reactant with an active site glutamate (Glu-30) residue.[162] This interaction is diminished as the hydride transfers from the co-factor to the substrate.

Recently, some QM/MM modelling has been done on the influenza-B sialidase mechanism: Jourand investigated the enzyme catalysed hydrolysis of N-acetyl-α-D-neuraminyllactose (NANL), **13**.[163] For the work discussed in this section, the Glu-277, Tyr-406 and Asp-151 side chains[164] were included in the QM region. The substrate was also included but the N-acetyl and glyceryl side chains of **13** were treated with the MM potential. The galactose residue in the lactose leaving-group was also included in the QM region but excluded the $-CH_2OH$ side chain on the 5-position (which was treated using MM). The number of atoms allowed to optimise

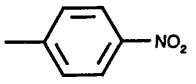
throughout the simulation was 407, of which 79 were treated with the semi-empirical AM1 Hamiltonian. The "flexible" atoms included protein residues within 6Å, and water molecules within 11Å of the substrate-C2 centre. The remaining atoms in the simulation (*ca.* 6400) were fixed.[163]



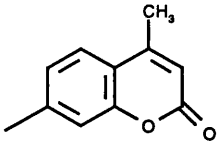
13 R=



14 R=

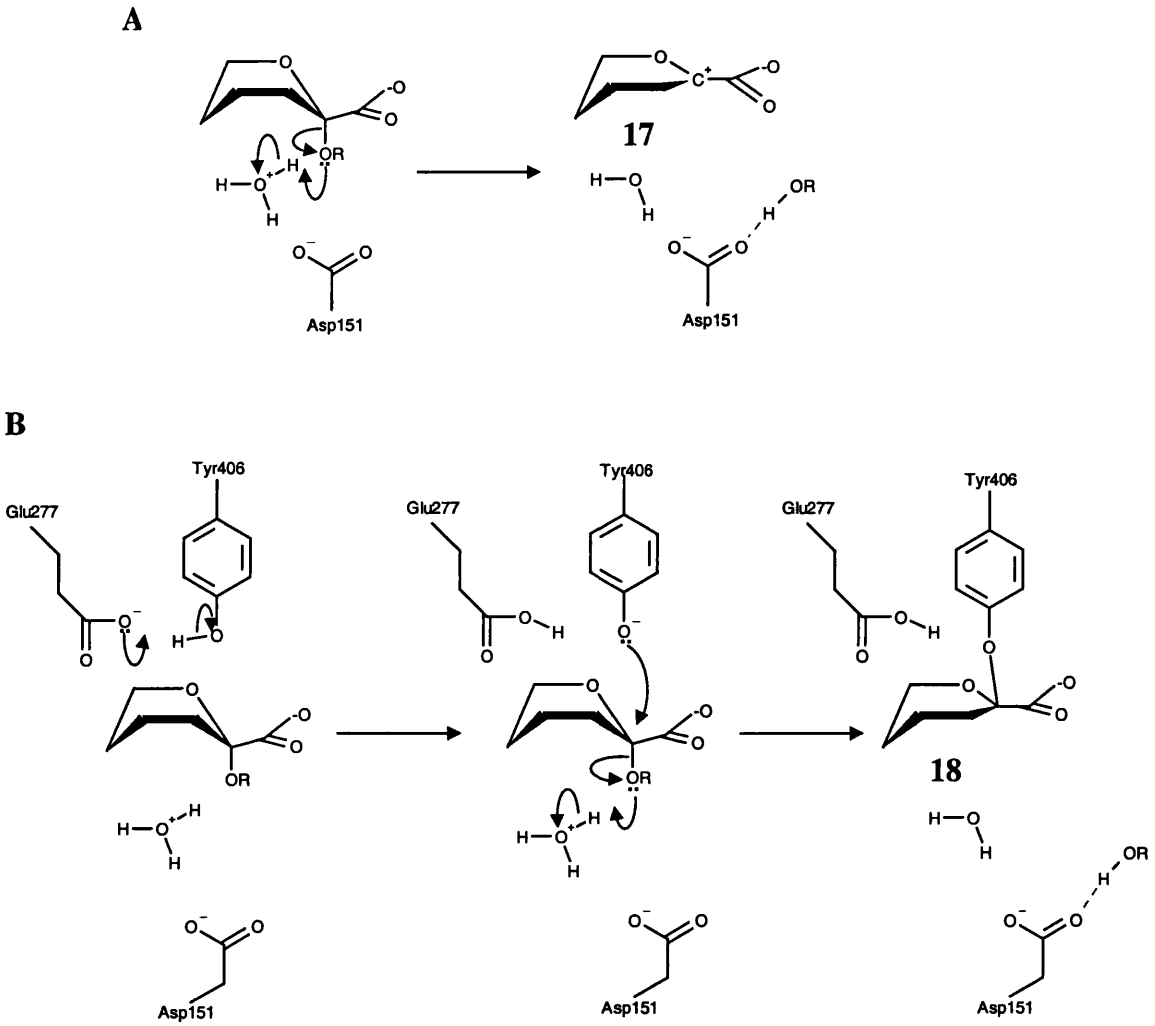


15 R=



16 R= –H

Jourand considered an acid catalysed mechanism involving a hydronium ion (scheme 8.11a), followed by nucleophilic addition of a water molecule to the resulting sialyl zwitterion,[163] **17**. Proton abstraction by Asp-151 was then considered as a subsequent step for generating the bound N-acetyl-α-D-neuraminic acid (Neu5ac, **16**) product. Jourand also considered the possibility of an alternative mechanism: the formation of a covalent intermediate involving deprotonated Tyr-406, **18**. Such an intermediate was found to lie within a potential energy well, being some 200 kJ mol^{-1} lower in energy than the sialyl zwitterion. Furthermore, in going from the bound reactant-state to the covalent intermediate (as shown in scheme 8.11b), the system was found to drop by about 240 kJ mol^{-1}.



Scheme 8.11

Further to Jourand's studies, Barnes and Williams adopted a QM/MM methodology to investigate the hydrolysis of *p*-nitrophenyl-N-acetyl-α-D-neuraminic acid (Neu5acPNP, **14**) by influenza-B sialidase.[165,166] They considered an energy surface describing the removal of the *p*-nitrophenolate leaving-group from the substrate, in concert with nucleophilic water attack at the positively charged C2 centre. This system included protein residues within 17Å of the substrate. A harmonic constraint was applied to protein atoms beyond 15Å of the Neu5acPNP and a stochastic boundary was applied to the solvent molecules. Some 65 atoms were included in the QM region which included the substrate (without the glyceryl or N-acetyl groups) and the Glu-277, Tyr-406 and Asp-151 side-chains. The N-acetyl and glyceryl groups of the Neu5acPNP were incorporated into the MM region.[165,166]

The energy surface was mapped out in 0.2Å steps and an approximate TS structure was located using a finer scan of 0.02Å around the "coarse" saddle-point. The Pauling bond-order in the TS structure for the cleaved glycoside bond was 0.00. Further, the bond order for the forming bond between the nucleophilic water and the substrate-C2 position was also very small (0.02). Therefore, this TS structure was representative of a "sialyl zwitterion like" species in close contact with the nucleophilic water and *p*-nitrophenolate leaving-group.

The resulting KIEs obtained from the TS are given in Table 8.1. Good agreement was found between the calculated ^{18}O-leaving-group KIE and the resulting experimental value.[2] However, the 2°-β-deuterium KIEs were inconsistent with the experimentally determined values.[2] The normal mode for the reaction co-ordinate was associated with a negative eigenvalue, but this was accompanied by several others. The normal modes associated with these negative eigenvalues were related to out-of-plane bends of the *p*-nitrophenolate aromatic ring. It should be noted that the TS was an approximate structure and that no functionality was available at that time to allow refinement. Furthermore, the experimental results referred to the catalysed reaction involving influenza-A sialidase and not the influenza-B type.

Table 8.1. Barnes and Williams' calculated Kinetic Isotope Effects (KIEs) for sialidase catalysed hydrolysis of **14**. The TS relates to a concerted step involving the removal of PNP with nucleophilic attack by water.

Effect	Calculated	V/K_m (Expt)
Pro-R 3-[^{2}H]	1.024	1.035±0.010
Pro-S 3-[^{2}H]	1.142	1.058±0.006
3-[^{2}H]$_2$	1.177	1.087±0.015
2-[^{18}O]	1.047	1.045±0.021

Pro-R and *pro-S* hydrons are defined above for structures 13-16; 3-[^{2}H]$_2$-both hydrogen atoms at the 3-position are deuterated; 2-[^{18}O]-glycoside oxygen isotope effect.

From the discussion presented in this section, it is evident that many approaches are available in modelling enzyme catalysed reactions using a hybrid QM/MM potential. Firstly, *ab initio* levels of theory can be used to describe the reacting system within a

rigid MM environment.[151,152] Secondly, a semi-empirical QM Hamiltonian can be coupled with a much larger and/or more flexible MM environment that neglects protein atoms beyond a certain distance of the active site.[153-155,159,161,162,165,166] Thirdly, a semi-empirical QM Hamiltonian can be coupled with a less flexible MM environment that includes all the atoms within the protein sub-unit.[156,157,26,160,163] The first and third approaches are suitable for "rigid" enzymes, but the first model neglects any long-range electrostatic interactions in the enzyme-substrate complex. The second model is more applicable for flexible enzymes involving larger molecular displacements throughout the course of the reaction, but is at risk of neglecting some longer-range interactions. Finally, TS searching algorithms are now being used in conjunction with a QM/MM potential to verify and characterise TS structures. This gives the possibility of providing further insight into enzyme catalysed mechanisms via calculation of barrier heights and other properties such as Kinetic Isotope Effects (KIEs).

However, unlike most of the investigations discussed in this section, the work of Barnes and Williams[165] allowed the interplay between theory and experiment through the calculation and comparison with experimentally determined KIEs.[2] However, such interplay is often difficult to achieve due to the lack and/or inaccessibility of experimental results. For the sialidase mechanism, however, experimental KIEs have been published in which the chemical step relates (in part) to the kinetically important steps.[2] Thus, this provides for an ideal theoretical study: one that has a window to "real-life" chemistry.

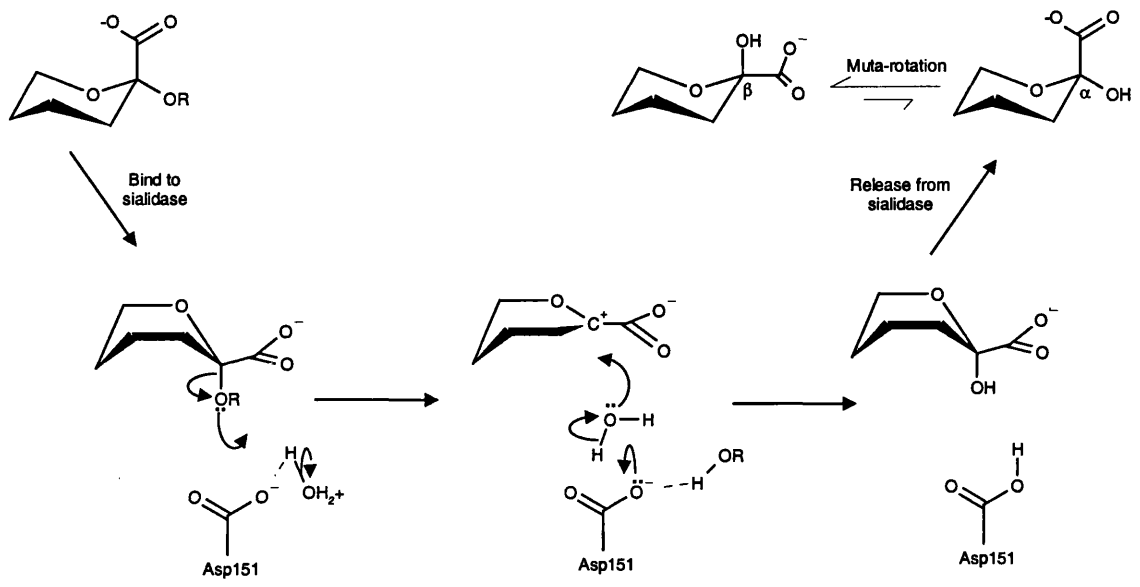
8.1.2. Influenza Sialidase Mechanism

From simple NMR studies, Chong *et al.* have provided evidence that influenza sialidase acts with retention of configuration, with the α-anomer product (N-acetyl-neuraminic acid, Neu5ac, **16**) initially produced before it undergoes muta-rotation to the more stable β-form.[167] In addition, they also calculated di-substituted 2^{o}-β-deuterium KIEs (3-$[^{2}H]_2$) for the sialidase catalysed hydrolysis of 4-methylumbelliferyl-N-acetyl-α-D-neuraminic acid (Neu5acUMB, **15**). The value for the effect on V_{max}/K_m at pH 6.0 was normal at 1.09 ± 0.02. However, in contrast, an

inverse effect on V_{max} of 0.976 ± 0.005 was also obtained.[167] The V_{max}/K_m effects suggested a degree of positive charge development at the substrate-C2 centre in the TS structure, resulting from the partially rate-limiting cleavage of the glycoside bond.

The effect on V_{max} represents a product of the KIEs for each isotopically sensitive step. For this reaction, these steps are likely to be the formation and destruction of the enzyme-intermediate complex. The inverse effect obtained at pH 6.0 led Chong *et al.* to conclude that breakdown of the intermediate was more rate limiting than its production.[167] They went on to suggest that an inverse isotope effect was consistent with the formation of a sialyl zwitterion intermediate, **17**, the KIE resulting from reduced hyperconjugation in the TS for the destruction of **17**. The inverse effect was inconsistent for the breakdown of the covalent intermediate, **18**, since the amount of hyperconjugation in the TS would be expected to increase or remain unchanged.

Chong *et al.* also noted a greater increase with pH in the 3-$[^2H]_2$ effect on V_{max}/K_m when using D_2O solvent.[167] From this result, they went on to suggest that formation of the sialyl zwitterion involved some form of proton transfer. Furthermore, they also suggested that "formation of the transition-state requires proton donation from solvent, facilitated by the negatively charged Asp-151". Thus, the mechanism may first involve general acid catalysis, leading to the formation of the sialyl zwitterion via cleavage of the glycoside bond (scheme 8.12).



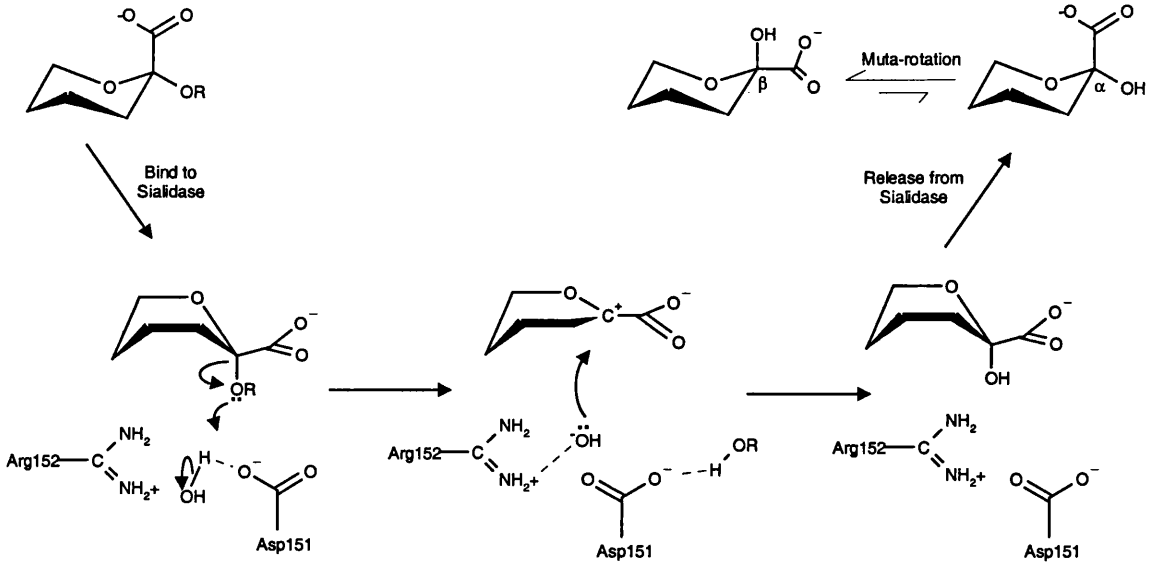
Scheme 8.12

Molecular dynamics studies also revealed the possibility of hydrogen bonding between a water molecule and active site residue, Arg-152, prior to cleavage of the glycoside bond.[30] The formation of a hydroxide anion in the proximity of deprotonated Asp-151 (and protonated Arg-152) results on proton transfer from the water to the glycoside oxygen atom in the substrate. Such a mechanism does not involve any general acid catalysis, and results in a mechanistic scheme that excludes direct involvement of any enzyme residue (scheme 8.13).[30] However, the formation of two anions in close proximity within the enzyme active site is likely to be unfavourable.

Guo *et al.* subsequently investigated the KIEs (experimental, table 8.1) for sialidase catalysed hydrolysis of *p*-nitrophenyl-N-acetyl-α-D-neuraminic acid (Neu5acPNP, **14**).[2] They discovered that the individual 3-[^{2}H] deuterium effects on V_{max}/K_m are not identical: the *pro-S* effect was larger than the *pro-R* effect. Guo *et al.* concluded from these results that the reacting conformation of the substrate was the $B_{2,5}$ conformation.[2] They subsequently argued that a TS structure closely resembling a sialyl zwitterion half-chair conformation would result in a larger *pro-R* effect than *pro-S*. This is due to the former having greater overlap with the empty p-orbital on the adjacent C2 centre in this conformation. By setting up simultaneous equations for the 2°-β-deuterium KIE, Guo *et al.* predicted the dihedral angles between the *pro-S*

and *pro-R* hydrons with the developing empty p-orbital at C2 to be 23° and 143° respectively.[2] For the Barnes and Williams TS structure,[165] the dihedral angles between the empty p-orbital on the C2 carbon and the β-hydrons were calculated to be 168° (*pro-R*) and 52° (*pro-S*).

The normal ^{18}O effect on V_{max}/K_m of 1.045 provided some evidence that the rate-determining step at pH 6.0 involves the cleavage of the glycoside bond. The calculated β_{lg} (V_{max}/K_m) value of –0.4 is significantly smaller than the typical value of –1, the latter usually obtained for both enzymic and non-enzymic reactions. Guo *et al.* suggested from this result that a degree of acid catalysis was being applied to the glycoside oxygen.

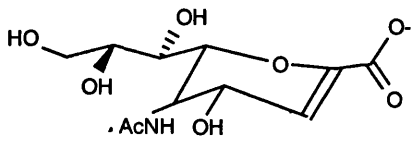


Scheme 8.13

The 3-[^2H]$_2$ effect on V_{max}/K_m for the isotopic substitutions was larger at pH 9.5, both for the Neu5acUMB[167] and Neu5acPNP[2] substrates. In the latter case, it was suggested that a kinetically silent step involving the *in situ* conformation change of the substrate from $^2C_5$ to $B_{2,5}$ was responsible for the suppression of the observed isotope effects at optimum pH.

For the Neu5acPNP substrate, the effect on V_{max} was near unity for all calculated isotope effects. This led Guo *et al.* to conclude that effects on V_{max} and V_{max}/K_m were associated with different processes, and suggested that no effect on V_{max} represented an isotopically silent, non-covalent event.[2] The normal solvent isotope effect of 1.9 of Chong *et al.* provides evidence for the breaking of hydrogen bonds in this step. It was suggested that such a step might involve a conformation change of the bound product from $B_{2,5}$ to $^2C_5$ prior to its release from the enzyme. The absent 3-$[^2H]_2$ normal effect on V_{max} obtained for Neu5acPNP[2] was in contrast to the inverse effect calculated by Chong *et al.* for Neu5acUMB.[167] Thus, the sole evidence for the formation of a sialyl cation was absent in the reaction involving Neu5acPNP. However, further work by the von Itzstein team confirmed their original finding of the inverse 3-$[^2H]_2$ effect on V_{max}.[168]

Crystallographic studies by Janakiraman *et al.* discovered that on soaking NA crystals with N-acetylneuraminyllactose (NANL), an entity that lacked the glycoside oxygen was bound to the enzyme active site.[169] This substrate was confirmed as not being the inhibitor, 2-deoxy-2,3-didehydro-N-acetylneuraminic acid (Neu5ac2en, **19**) since the electron density map was not consistent with this structure. Janakiraman *et al.* suggested that this species was the sialyl zwitterion,[169] **17**. They went on to suggest that the role of the active site residues (Glu277, Tyr406 and Asp151) is not only to stabilise the positive charge developed in the TS, but to also maintain planarity about the C2 centre. Furthermore, it was suggested that the driving force for the reaction comes from the induction and stabilisation of **17** and not from proton transfer or nucleophilic attack. This implies the purpose of sialidase is purely structural: to stabilise a "sialyl zwitterion like" structure. However, it should be noted that the results based on these postulations report an equilibrium distance for the C2—O_h bond as 1.41Å for the bound sialyl zwitterion. This is indicative of a C—O single bond and not a partial double bond that is expected in **17** (*cf.* C2—O_h in Neu5ac2en has an equilibrium length of 1.37Å). Considering the short lifetime of unbound sialyl zwitterion, it is somewhat surprising to see it isolated in the crystal structure while bound to the active site.



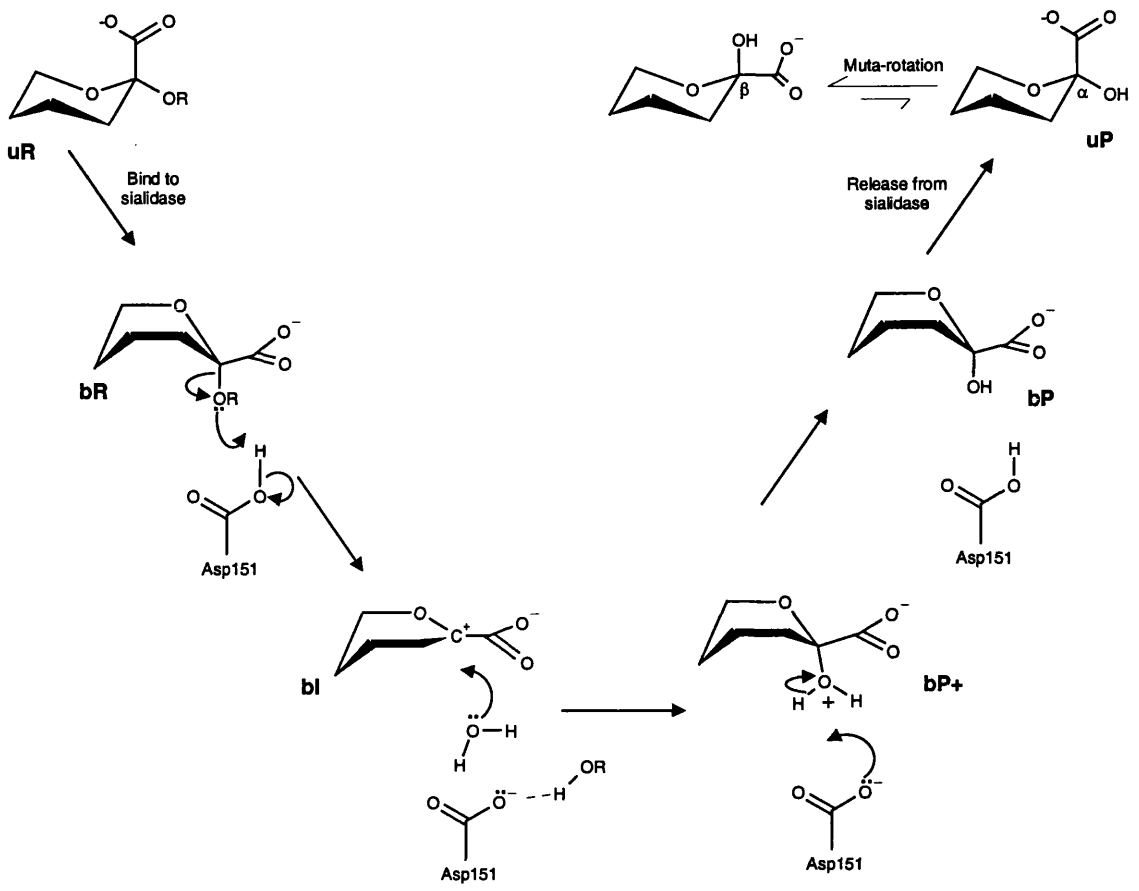
19

Interestingly, Smith has suggested from the results of *ab initio* investigations on 2-oxonol, that sialidase weakens the glycoside linkage by binding the substrate in the boat conformation.[170] Smith went onto suggest that general acid catalysis involving Asp(H)-151 would result in a short C2—O_g distance in the TS. It was suggested that this residue could also act as a base, facilitating the removal of a proton from an attacking water molecule on addition to the zwitterion intermediate.[170]

The findings and discussion given in this section seem to be consistent with the mechanism shown in Scheme 8.14. The first step involves binding the unbound reactant (uR) to the sialidase enzyme in the $B_{2,5}$ conformation. The bound substrate (bR) is then subjected to general acid catalysis, resulting in the cleavage of the glycoside bond. In principle, this could be provided (at optimum pH) by the Asp(H)-151 residue. The general acid catalysis step results in the formation of an enzyme-bound sialyl zwitterion intermediate (bI), which is subsequently destroyed by nucleophilic attack of water at the positively charged centre. The deprotonated Asp-151 then abstracts a proton from the protonated product (bP^+) to give the bound product (bP). This is then released from sialidase as the α-anomer (uP) and is then converted to the β-anomer through muta-rotation.

In this chapter, the aim is to present further insight into the mechanisms of influenza sialidase. The work concentrates on two major steps associated with the mechanism proposed in Scheme 8.14. Firstly, the general acid catalysis by Asp(H)-151 on Neu5acPNP and secondly, base catalysis by Asp-151 on the protonated product. The latter step follows the attack of water to the sialyl zwitterion intermediate.

The AM1 Hamiltonian reproducibility of gas phase experimental and high level *ab initio* results are considered first, followed by investigation and analysis into both proton transfer steps in scheme 8.14. This is followed by a brief investigation into the attack of water on the bound sialyl zwitterion. KIEs associated with both proton transfer steps are then calculated and discussed. Finally, the flexibility of the system in these simulations is considered and is used to provide some knowledge on how to set up future simulations. This is followed by a conclusion.



Scheme 8.14

8.2. Performance and Reliability of the AM1 Hamiltonian

The majority of work contained with this chapter is concerned with proton transfers. The energy of such processes in the gas phase can be evaluated directly from individual proton affinities determined *in vacuo*. Proton affinity of a base is defined as the energy required to remove a proton form the corresponding acid, i.e.

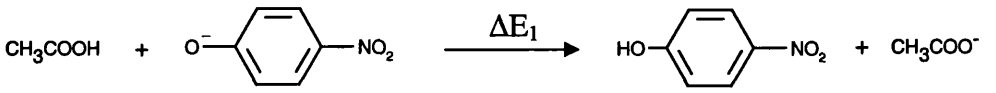
$$BH \longrightarrow B^- + H^+ \qquad (8.1)$$

Dewar and Dieter have shown that AM1 reproduces experimental proton affinities well.[171] Various calculated proton affinities and experimental values are given in Table 8.2. These processes relate to the proton transfer steps in scheme 8.14.

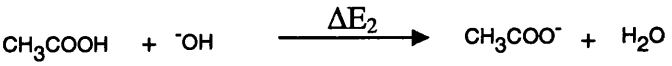
Table 8.2. Calculated and experimental proton affinities (in kJ mol^{-1}) for conjugate bases of interest to both proton transfer steps in scheme (8.14).

BH	B^-	ΔH_{PA} (AM1)	ΔH_{PA} (Expt)
$NO_2C_6H_4$—OH	$NO_2C_6H_4$—O^-	1346.0[a]	1345.2[a]
CH_3COOH	CH_3COO^-	1482.0[b]	1458.1[b]
H_2O	HO^-	1725.1[b]	1635.1[b]

[a] Reference 172. [b] reference 171



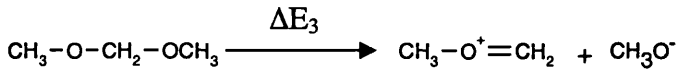
(8.2)



(8.3)

It is evident from the AM1 calculated proton affinities that ΔE_1 in eqn. 8.2 is over-estimated by about 23 kJ mol^{-1}. Furthermore, the proton transfer between acetic acid and hydroxide anion (scheme 8.3), is under-estimated by about 66 kJ mol^{-1}. Therefore, using the Hammond postulate, the proton transfer from acetic acid to phenolate will be too advanced in the TS when using the AM1 Hamiltonian. In contrast, the proton transfer from acetic acid to the hydroxide species will be too early when using the AM1 method.

The error associated with the cleavage of the glycoside bond is more difficult to quantify. Oxocarbenium ions have very limited stability and consequently experimental thermochemical data is difficult to find. To provide an approximate value for the AM1 Hamiltonian error associated with this process, the energy change, ΔE_3, for the dissociation of dimethoxymethane into the resulting oxocarbenium ion and methoxide anion was evaluated (eqn. 8.4).



(8.4)

For this purpose, *ab initio* calculations at the MP2/6-31+G(d) level of theory were carried out for all species shown in eqn. 8.4, the results are given in table 8.3.

Table 8.3. Energies for all species given in eqn. 8.4. AM1 energies are heats of formation in kJ mol^{-1}; MP2/6-31+G(d) energies are given in Hartrees (E_h). Zero-point vibrational energies (ZPE) calculated at MP2/6-31+G(d) are also listed.

SPECIES	AM1/kJ mol^{-1}	MP2/6-31+G(d)/E_h	ZPE/kJ mol^{-1}
$CH_3OCH_2OCH_3$	-432.18	-268.71788	305.26
$CH_3OCH_2^+$	661.05	-153.62500	96.46
CH_3O^-	-161.06	-114.74454	184.94

Thus, ΔE_3 in equation 8.4 is evaluated as 932.2 kJ mol^{-1} using the AM1 Hamiltonian:

$$\Delta E_3 = -\Delta H_f(CH_3OCH_2OCH_3) + \Delta H_f(CH_3OCH_2^+) + \Delta H_f(CH_3O^-)$$

$$\Delta E_3 = -(-432.18) + 661.05 + (-161.06)$$

$$\Delta E_3 = 932.2 \text{ kJ mol}^{-1}$$

At MP2/6-31+G(d), ΔE_3 is calculated as:

$$\Delta E_3 = -E_h(CH_3OCH_2OCH_3) + E_h(CH_3OCH_2^+) + E_h(CH_3O^-)$$

$$\Delta E_3 = -(-268.71788) + (-153.62500) + (-114.74454)$$

$$\Delta E_3 = 0.34834\ E_h \text{ (1 Hartree, } E_h = 2625 \text{ kJ mol}^{-1})$$

$$\Delta E_3 = 914.47 \text{ kJ mol}^{-1}$$

The enthalpy change at absolute zero can be evaluated by accounting for the ZPE of all species in eqn. 8.4:

$$\Delta E_3' = \Delta E_3 - \Delta ZPE$$

$$\Delta E_3' = 914.47 - (305.26 - 96.46 - 184.94)$$

$$\Delta E_3' = 890.7 \text{ kJ mol}^{-1}$$

Thus, the calculated enthalpy change (obtained after applying unscaled ZPE corrections) at the MP2/6-31+G(d) level of theory is 890.7 kJ mol^{-1}. Thermal corrections are likely to be small and are therefore not included here. BSSE effects must now be evaluated since they could significantly affect the value of ΔE_3. The energies of both fragments with their super-molecule geometries (E_{frag}) and in the presence of counter-fragment ghost orbitals (ε_{frag}) are given in table 8.4.

Table 8.4. Fragment energies calculated at MP2/6-31+G(d), with their "super-molecule geometry", E_{frag} and in the presence of ghost orbitals of the counter-fragment, ε_{frag}. These are used to calculate the BSSE correction associated with the dissociation process in eqn. 8.4

Fragment	E_{frag}	ε_{frag}
$CH_3OCH_2^+$	-153.55185	-153.55664
CH_3O^-	-114.73814	-114.74812

The BSSE effect at MP2/6-31+G(d) for the dissociation process shown in eqn. 8.4 can be calculated as:

$$BSSE = \varepsilon_{frag} - E_{frag}$$

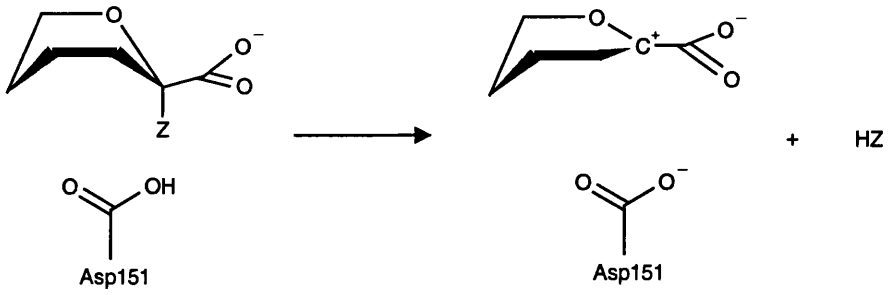
$$BSSE = [(-153.55664) + (-114.74812)] - [(-153.55185) + (-114.73814)]$$

$$BSSE = -0.01477\ E_h = -38.8\ kJ\ mol^{-1}$$

Thus, the effect of BSSE at MP2/6-31+G(d) is to decrease ΔE_3 by 38.8 kJ mol^{-1} to 851.9 kJ mol^{-1}. Therefore, the ΔE_3(AM1)-ΔE_3(MP2/6-31+G(d)) difference is evaluated as 80.3 kJ mol^{-1} after ZPE and BSSE corrections have been applied to the *ab initio* result. However, although thermal and scaling corrections have not been applied, this value provides some evidence that AM1 may be over-estimating the difficulty of removing a leaving group to form the oxocarbenium ion. Thus, the effects discussed in this section might lead to some suppression in the calculated KIEs for the full sialoside/enzyme system.

8.3. Proton Transfer Steps from Asp(H)-151 to Sialidase-Bound Substrate

Scheme 8.14 includes two proton transfer steps involving Asp-151 as an acid (in the protonation of Neu5acPNP) and as a base (in abstracting a proton from the protonated product). In the following discussion, both processes are considered as general acid catalysis steps (scheme 8.15, Z=$OC_6H_4NO_2$ or Z=OH).



Scheme 8.15

For both proton transfers, two initial structures were generated in which the "transferring proton" (H_p) was either bonded to the Asp-151 carboxyl group (R_{start}), or attached to the substrate glycoside oxygen, O_g (P_{start}). Some selected geometrical parameters from the optimised starting structures are given in table 8.5. From R_{start}, H_p was adiabatically mapped to its cartesian co-ordinate in P_{start} using 0.1Å steps. For the simulation involving the Neu5acPNP substrate, each structure on the proton transfer (PT) profile was optimised for 600 cycles. Due to CPU time constraints, each structure on the corresponding profile for PT to Neu5ac was subjected to only 200 cycles of optimisation before H_p was moved. Both profiles were generated in the same way: by fixing the Asp-COO—H_p distance at each stage of the PT process.

Table 8.5. Various distances found in the initial reactant (R_{start}) and product (P_{start}) structures for both substrates. The position of proton, H_p in R_{start} and P_{start} defines the vector to be mapped. C2—O_g is the distance between the anomeric carbon and glycoside oxygen; COO—H_p is the distance between initially protonated carboxyl oxygen and H_p. H_p—O_g is the distance between the transferring proton and glycoside oxygen.

	Z=$OC_6H_4NO_2$		Z=OH	
	R_{start}	P_{start}	R_{start}	P_{start}
C_2—O_g	1.460	2.718	1.421	1.545
COO—H_p	0.978	2.063	0.979	2.593
H_p—O_g	2.031	0.977	2.094	1.010

The coarse energy profile (figure 8.1a) for the proton transfer to Neu5acPNP results in spontaneous cleavage of the glycoside bond with a corresponding sharp drop in the total energy. A "true" energy profile would involve a smooth energy (and structural) change in going from reactants and products. Thus, it is unlikely that the resulting barrier heights from this profile have any significance. The proton transfer to Neu5ac also results in a coarse, discontinuous energy profile (figure 8.1b). The comparison of all structures on the reaction co-ordinate with R_{start} gives direct evidence of a discontinuity in both PT steps. A profile involving conservative changes in structure would yield a straight, diagonal line on the RMS deviation plot. These structural comparisons involve a subset of atoms, namely the Asp-151 and substrate. Therefore, the RMS deviation plots given in figure 8.1(c)-(d) do not correspond exactly with the total energy profiles in figures 8.1(a)-(b).

The pseudo-maximum on both profiles (indicated with an arrow in figure 8.1) was used as an initial guess for the TS structure corresponding to the relative PT step studied. For the proton transfer to Neu5acPNP, the pseudo-maximum had a COO—H distance of 1.426Å, and was located immediately before the spontaneous cleavage of the glycoside bond. For transfer of the proton to Neu5ac, the pseudo-maximum was obtained at a COO—H distance of 1.357Å. These structures were refined (section 3.3.1) and the IRC was followed from them to the connecting reactant and product wells. For these calculations, the CORE was defined as the QM region without the

QM link atom (which was incorporated into the ENVIRONMENT). Herein, this is referred to as the "small core".

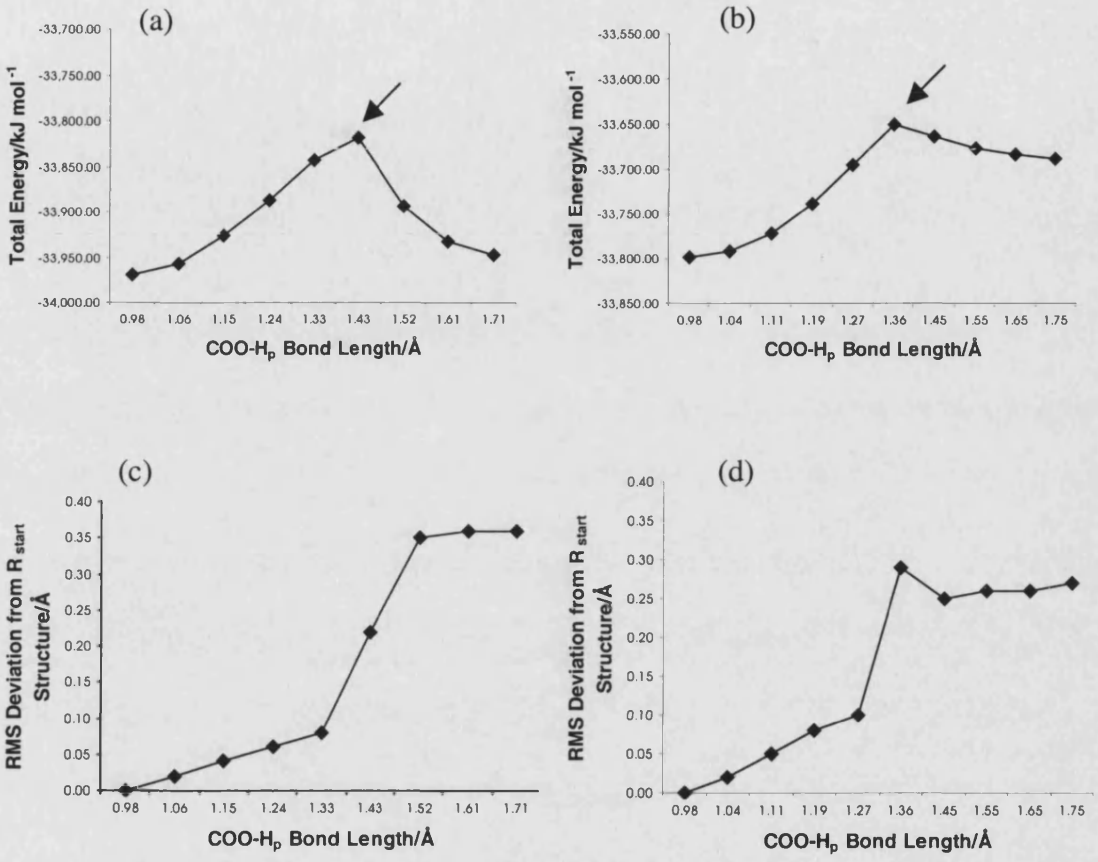


Fig 8.1. Energy profiles for proton transfer steps from Asp(H)-151 to a) Neu5acPNP and b) Neu5ac. RMS deviation from R_{start} of structures on the coarse reaction profile for c) PT to Neu5acPNP and d) PT to Neu5ac. RMS deviation comparisons were made using the substrate and Asp(H)-151 residue. Arrows in a) and b) point towards the unrefined approximate TS.

The refined TS for acid catalysis applied to Neu5acPNP has an imaginary frequency of $653i$ cm^{-1}. The corresponding normal mode for this frequency involves proton transfer from Asp(H)-151, in concert with the glycoside bond stretch in Neu5acPNP. The imaginary frequency in the TS associated with the proton transfer to Neu5ac ($517i$ cm^{-1}) has a corresponding normal mode involving PT allied with slight planarisation about the C2 centre of the substrate.

In both transition structures, proton donation is advanced. The H_p—O_g distance for Neu5acPNP and Neu5ac is 1.092Å and 1.076Å respectively. Thus, in the case of Neu5ac, a greater amount of acid catalysis is being applied in the TS. However, for

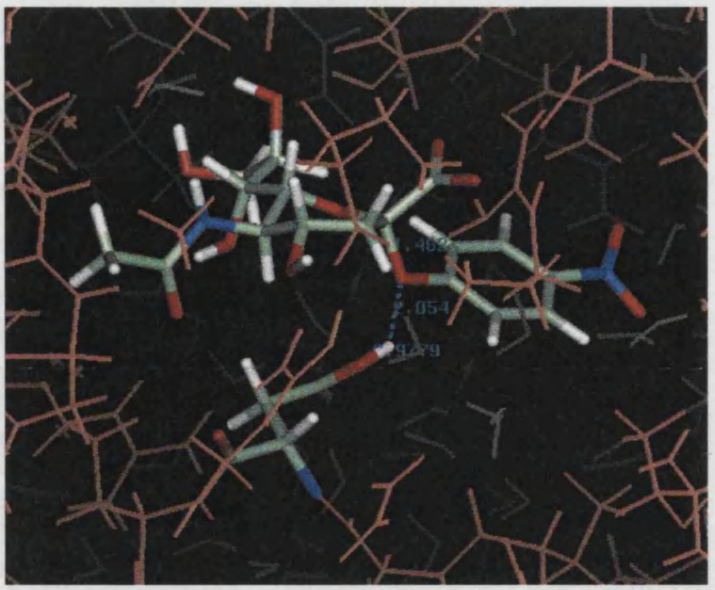
Neu5acPNP, it was noted in section 8.2 that the energy required to cleave the glycon-aglycon bond may be over-estimated by the AM1 Hamiltonian. This may be alleviated in part by the more advanced acid catalysis in the TS found at this level of theory (section 8.2).

For the proton transfer to Neu5acPNP, there is elongation of the glycoside bond in the TS (1.610Å) over the reactant-state (1.462Å). The resulting sialyl zwitterion has a C2—O_g distance of 2.684Å, with a hydrogen bond between the deprotonated Asp-151 and the *p*-nitrophenol hydroxyl group (AspO—H_p distance of 2.031Å). In going from reactant to sialyl zwitterion through the TS, there is planarisation of the C2 centre with concomitant decrease in the C2—C3 and C2—O_h distances (which is evidence of stabilisation through conjugation and hyperconjugation respectively). These effects are also seen in protonation of the Neu5ac substrate, but to a lesser extent. The reactant, TS and product structures for both processes are shown in figure 8.2, indicating the distances for the glycoside bond and those involving the transferring proton. Table 8.6 lists selected geometrical parameters for the reactant, TS and product structure for both proton transfer steps.

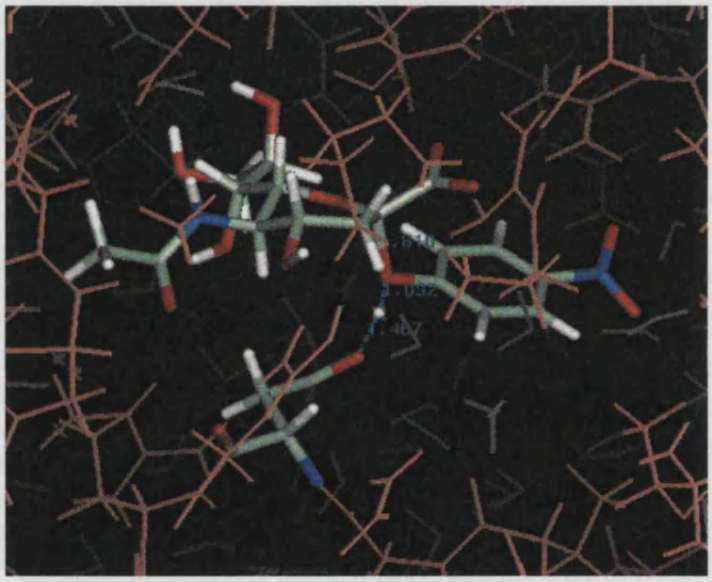
The total energy of the reactants, TS and products can be given as the sum of the individual contributions by the QM, QM/MM and MM potentials, i.e.:

$$E_{TOT} = E_{QM} + E_{int} + E_{MM} \tag{8.5}$$

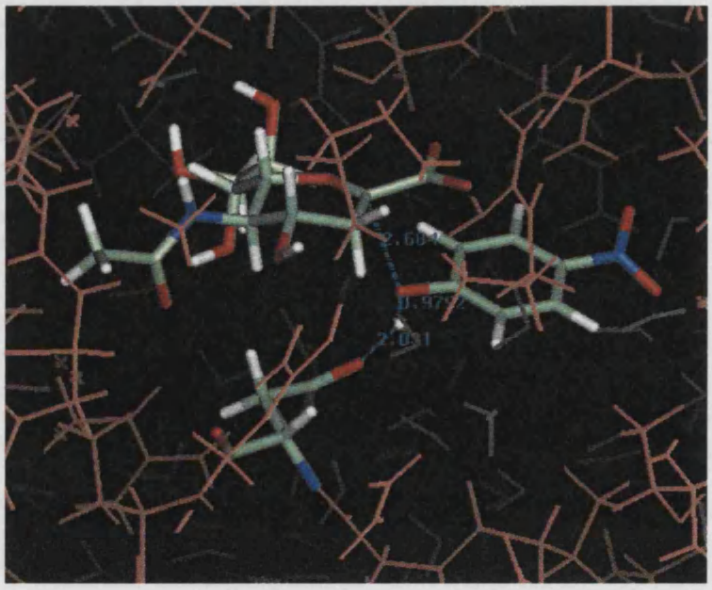
where, E_{QM} the energy of the QM system, E_{int} is the interaction energy between the QM and MM regions and E_{MM} is the energy of the MM system. Table 8.7 gives these individual contributions for each stationary point associated with both PT processes.



bR

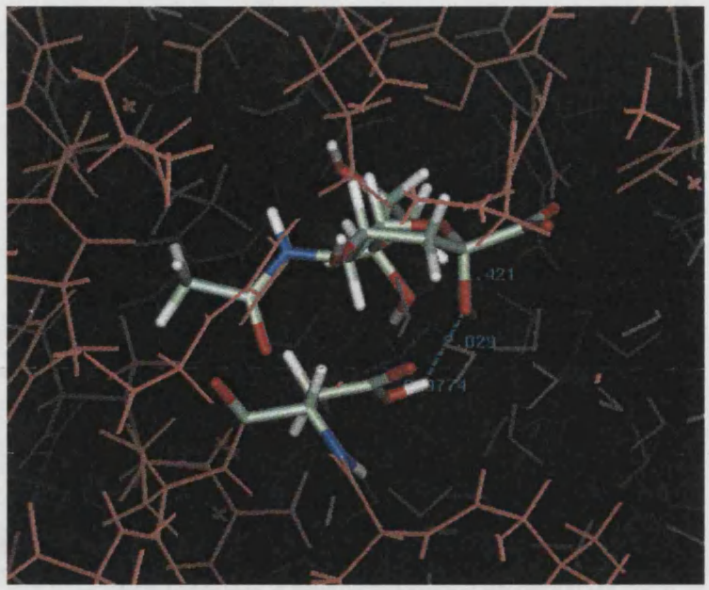


TS1

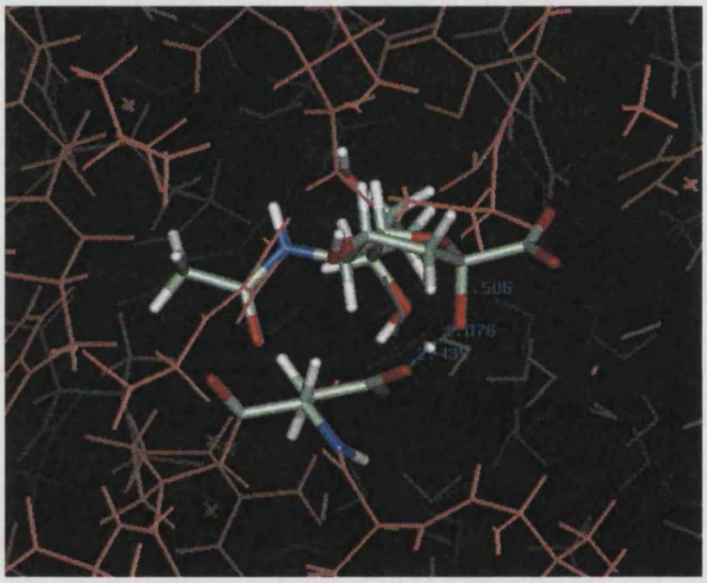


bI

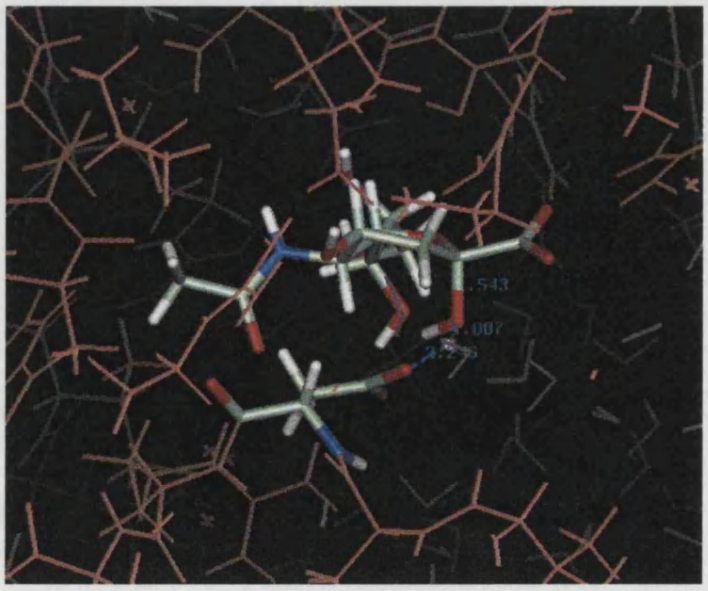
Fig 8.2a. The reactant (bR), transition state (TS1) and product (bI) for general acid catalysis applied to Neu5acPNP by Asp-151 of influenza-A sialidase. Asp-151 and substrate are shown using the "liquorice" solid model, the remaining enzyme is coloured brown and solvent is coloured grey. Distances are in Å.



bP



TS2



bP$^+$

Fig 8.2b. The reactant (bP), transition state (TS2) and product (bP$^-$) for general acid catalysis applied to Neu5ac by Asp-151 of influenza-A sialidase. Asp-151 and substrate are shown using the "liquorice" solid model, the remaining enzyme is coloured brown and solvent is coloured grey. Distances are in Å.

Table 8.6. Selected geometrical parameters for unbound reactant (uR), bound reactant (bR), bound intermediate (bI), bound protonated product (bP^+), bound product (bP) and unbound product (uP). Parameters are also given for both proton transfer transition structures, TS1 and TS2.

Structure	C_2-O_g	AspO-H_p	O_g-H_p	C_2-O_h	C_2-C_3	C_3-H_r	C_3-H_s	O_h-C_2-C_1	O_h-C_2-C_3	C_3-C_2-C_1	ΣC_2
uR	1.426	-	-	1.422	1.537	1.119	1.120	114.2	110.5	111.7	336.4
bR	1.462	0.978	2.054	1.408	1.526	1.125	1.124	112.3	118.8	109.9	341.0
TS1	1.610	1.467	1.092	1.364	1.504	1.127	1.126	110.8	121.9	113.7	346.4
bI	2.684	2.031	0.979	1.295	1.461	1.140	1.121	113.8	124.4	121.4	359.6
bP^+	1.543	1.937	1.007	1.368	1.509	1.129	1.118	109.5	119.5	113.7	342.7
TS2	1.506	1.524	1.076	1.379	1.510	1.128	1.118	108.0	119.2	112.5	339.7
bP	1.421	0.977	2.029	1.409	1.523	1.123	1.119	106.6	116.3	111.0	333.9
uP	1.411	-	-	1.427	1.526	1.120	1.119	114.2	112.0	112.0	338.2

Structure	O_g-H_r	O_g-H_s	O_g-C_2-C_3-H_r	O_g-C_2-C_3-H_s	O_h-C_2-C_3-C_4	O_h-C_6-C_5-C_4	C_6-O_h-C_2-C_3
uR	2.460	2.730	-42.2	75.9	-53.6	60.5	66.8
bR	2.336	3.213	17.5	132.3	4.1	69.3	18.1
TS1	2.427	3.353	22.2	137.5	4.6	62.5	6.7
bI	2.554	3.802	100.2	-143.5	-21.6	48.5	10.5
bP^+	2.518	3.362	31.5	146.8	20.0	55.1	-28.8
TS2	2.477	3.319	29.9	144.9	20.6	54.1	-30.4
bP	2.449	3.296	34.3	151.0	29.4	54.3	-37.6
uP	2.506	2.762	-43.7	74.5	-55.5	55.8	56.1

Table 8.7. Energies in kJ mol^{-1} for the reactant, TS and product structures associated with the proton transfers from Asp(H)-151 to Neu5acPNP ($Z=OC_6H_4NO_2$) or Neu5ac (Z=OH). E_{QM} is the energy of the QM system as defined by the QM Hamiltonian; E_{int} is the interaction energy between the QM and MM regions (defined by the QM/MM Hamiltonian); E_{MM} is the MM energy defined by the MM force field.

Z	Stationary Point	E_{QM}	E_{int}	E_{MM}
$-OC_6H_4NO_2$	Reactant	-2045.66	-2653.21	-29 283.81
	TS	-1837.69	-2774.74	-29 240.49
	Product	-1851.34	-2859.63	-29 211.61
—OH	Reactant	-2254.61	-2495.15	-29 115.91
	TS	-2102.84	-2549.48	-29 057.24
	Product	-2110.71	-2561.14	-29 064.84

For the proton transfer to Neu5acPNP, the activation barrier in the presence of the enzyme is 80 kJ mol^{-1} lower than the corresponding change *in vacuo* (figure 8.3.). Furthermore, the energy difference between the reactant-state and the sialyl zwitterion is reduced by 144 kJ mol^{-1} compared to the same change in the gas phase. Thus, the sialidase enzyme stabilises the TS and the resulting zwitterion as evidenced by the increasingly strong interactions between the QM and MM regions (E_{int}, table 8.7). Under enzymic conditions, the sialyl zwitterion formed on protonation of Neu5acPNP is only 50 kJ mol^{-1} less stable than the bound reactant.

For the proton transfer to Neu5ac within the enzyme-solvent environment, the energy difference between the protonated substrate and Neu5ac is reduced by about 15 kJ mol^{-1} compared to the same chemical change *in vacuo*. Within the enzyme environment, the protonated sialic acid is some 130 kJ mol^{-1} less stable than the Neu5ac.

These comparisons are not significant in terms of actual mechanistic conclusions since the barriers for the enzyme catalysed reaction should be compared to the those obtained in solution (and not in the gas phase). Furthermore, it should be emphasised that the *in vacuo* energy barriers discussed above refer to the change in the QM energy (E_{QM}) without the surrounding enzyme/solvent environment. Therefore, the

activation energies presented here would not correspond to the barriers obtained for the "unconstrained" gas phase reaction. However, the E_{QM} barriers reveal that sialidase does stabilise the TS associated with proton transfer to Neu5acPNP and the resulting zwitterion intermediate (with respect to the reactant). Sialidase may also afford some stabilisation to the protonated Neu5ac, with respect to the unprotonated sialic acid.

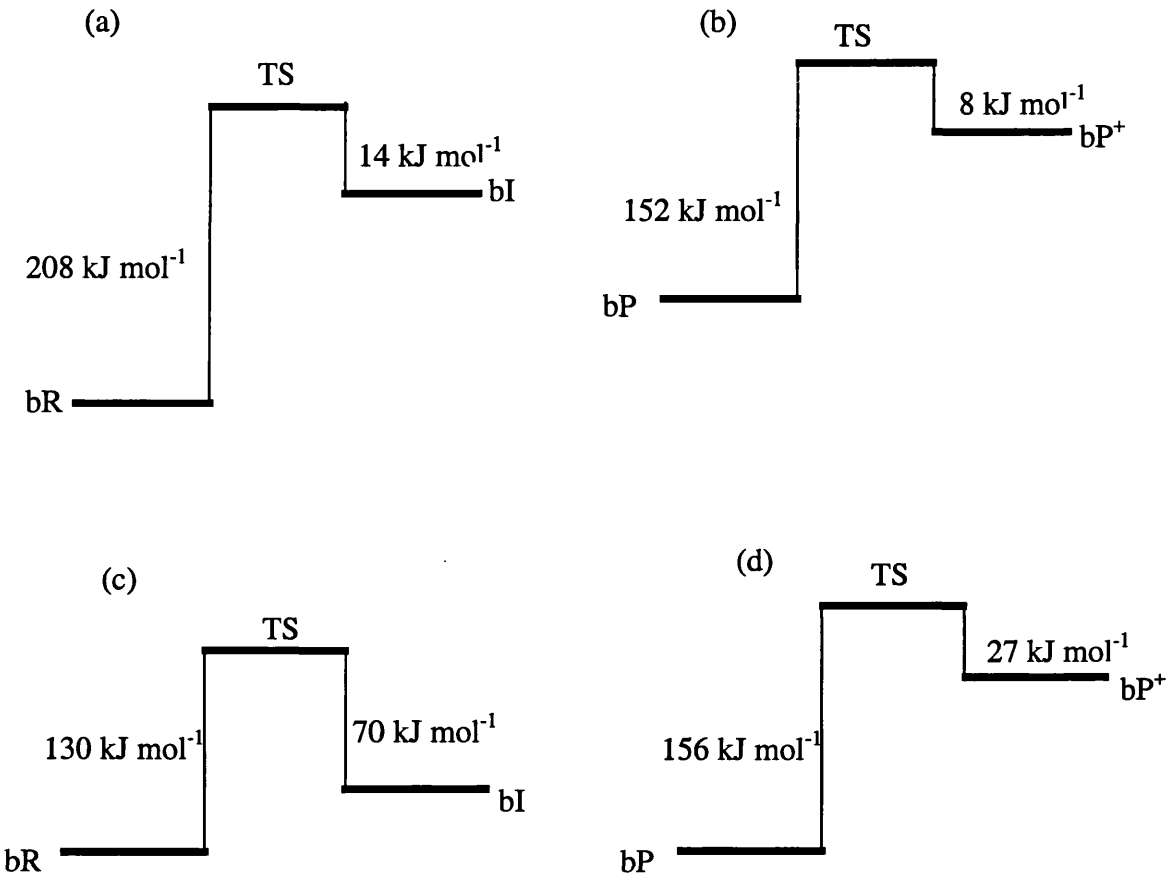


Fig 8.3. Energy barriers for a). proton transfer to Neu5acPNP *in vacuo*, b). proton transfer to Neu5ac *in vacuo*, c). Proton transfer to Neu5acPNP substrate in sialidase active site and d). proton transfer to Neu5ac in sialidase active site. "*in vacuo*" refers to the QM part of the ES system in the absence of the surrounding enzyme.

It has been reported that semi-empirical techniques over-estimate activation barriers for intermolecular proton transfer processes.[173] PM3 has been reported to over-estimate activation barriers for these processes by about 60-80 kJ mol^{-1} compared to *ab initio* studies.[173] Furthermore, AM1 was found to over-estimate these barriers by a "greater amount".[173] In addition to these limitations, this chapter has only considered one conformation of the protein/solvent environment. The work in chapter 6 demonstrated that the environment has a significant effect on the energetics

of a process. Thus, values reported in figure 8.3 are presented for completeness; they are only semi-quantitative at best.

The proton transfer steps discussed above may provide more insight into the mechanistic role of sialidase. While the —OH group is a conjugate base of a weak acid (water has a pK_a of about 14 at 25°C), *p*-nitrophenolate is a conjugate base of *p*-nitrophenol, a much stronger acid. The pK_a values[174] of the parent acids are given in table 8.8, along with the values of ethan-1,2-diol and propan-1,2,3-triol. The above enzyme-substrate simulations suggest that protonation of increasingly facile leaving-groups result in cleavage of the glycoside bond (*cf.* Neu5acPNP). For a poorer leaving group, protonation is required before cleavage of the C2—O_g bond takes place. The pK_a values for the diol and triol in table 8.7 are similar to that of water. The first residue in the "*in vivo*" leaving group is a galactose moiety, which is unlikely to be much more acidic than ethan-1,2-diol or propan-1,2,3-triol. Thus, extrapolating back from the simulations presented here, a lactose leaving-group may require protonation prior to cleavage of the glycoside bond.

Table 8.8. Experimental pK_a values for various acids at 25°c

Acid	pK_a
$HOC_6H_4NO_2$	7.15
H_2O	14.00
$CH_2(OH)CH_2(OH)$	14.22
$CH_2(OH)CH(OH)CH_2(OH)$	14.15

8.4. Nucleophilic Attack of Sialyl Zwitterion Intermediate by Water

The protonated Neu5ac can also be regarded as strong associated complex comprising of a sialyl zwitterion and a water molecule. Treating the associated water molecule with the TIP3P potential, the system was re-optimised until the RMS gradient fell below 0.02 kJ mol^{-1} $Å^{-1}$. This MM water was then treated with the AM1 Hamiltonian (along with the zwitterion and Asp-151) and the whole system was optimised further with the same gradient tolerance. The final distance between the

water molecule oxygen and the C2 centre was 4.090Å. Subsequently, the water molecule was translated back to its position in the associated complex in 0.1Å steps (allowing 600 cycles of ABNR optimisation for each structure).

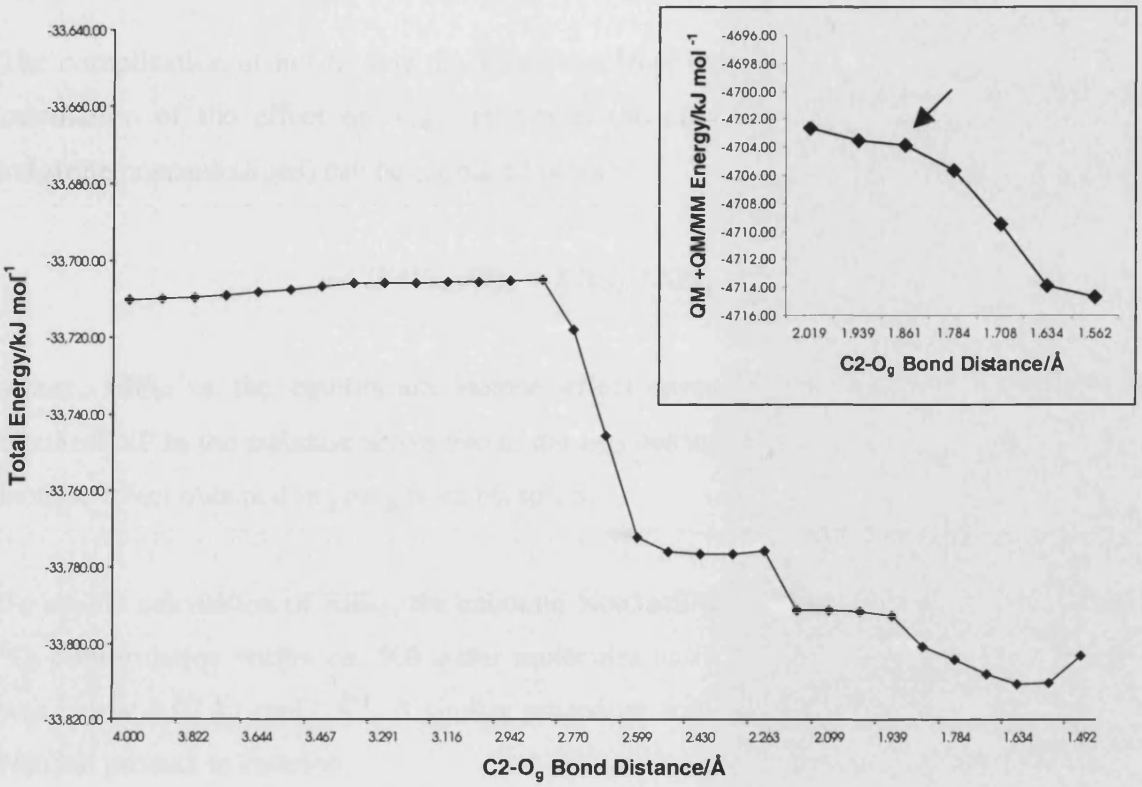


Fig 8.4. Total energy profile (kJ mol^{-1}) for the approach of a water molecule partaking in a nucleophilic attack on the sialyl zwitterion. The profile was mapped out using steps of 0.1Å. Inset shows the last part of the (QM+QM/MM) energy profile, with the arrow pointing towards a possible approximate TS.

Although the total energy (TE) profile is discontinuous, an initial test was carried out to determine whether a TS existed with a C2—O_g length of around 2.856Å (prior to the sharp drop in energy on the TE profile). The hessian matrix for this structure contained no negative eigenvalues. However, a smoother increase in the (QM + QM/MM) energy profile was obtained towards the end of the mapped co-ordinate. This suggested the presence of a maximum at *ca.* 1.86Å for the C2—OH_2 distance (figure 8.4, inset). Calculation of the hessian matrix and 16 steps of P-RFO TS refinement led to a structure with one negative eigenvalue associated with the C2—OH_2 bond stretch. However, the resulting RMS and MAX gradients of 2.8 and 12.8 kJ mol^{-1} $Å^{-1}$ were above the gradient tolerance levels of 0.4 kJ mol^{-1} $Å^{-1}$ and 1.2 kJ mol^{-1} $Å^{-1}$ respectively. Thus, this partially refined structure requires further

optimisation before being designated a TS structure; it has not been possible to do this due to CPU time restraints.

8.5. Kinetic Isotope Effects for Both Proton Transfer Steps

The complication of not finding the TS for nucleophilic water attack precludes the calculation of the effect on V_{max}. However, the effect on V_{max}/K_m (at limiting substrate concentrations) can be estimated using:

$$KIE\,(V_{max}/K_m) = EIE_{b1} * KIE_{pt1} \qquad (8.6)$$

where, EIE_{b1} is the equilibrium isotope effect associated with the binding of Neu5acPNP to the sialidase active site in the $B_{2,5}$ conformation. KIE_{pt1} is the kinetic isotope effect obtained in going from bR to TS1.

To enable calculation of EIE_{b1}, the unbound Neu5acPNP (uR) was optimised in the $^2C_5$-conformation within *ca.* 500 water molecules until the gradient for the system was below 0.02 kJ mol^{-1} $Å^{-1}$. A similar procedure was used to obtain the unbound Neu5ac product in solution.

The calculated isotope effects are given in table 8.9. The theoretical V_{max}/K_m results are in disagreement with the experimentally determined values. Firstly, the β-hydron effects are inverse and nearly equal. This suggests the TS structure conformation is incorrect, since the reacting boat conformation of the substrate would lead to different values for the *pro-R* and *pro-S* effects. As seen in figure 8.2a, the bound reactant obtained from TS1 is very flat about the C2 centre. The C6—O_h—C2—C3 dihedral angle of 18.1° is much smaller than that found in a full $B_{2,5}$ conformation (*ca.* 60°). Furthermore, this TS structure is also flat about the C2 centre (with a C6—O_h—C2—C3 dihedral of 7°). These results are thus not consistent with the experimental KIEs or with the bound substrate conformation in the crystal structure. A larger CORE (containing all protein residue side-chains within 6Å of the C2 reacting centre, as well as any water molecules within this region) was also used to evaluate the KIEs. However, little deviation in the effects were found when

Table 8.9. Isotope effects (2°-β-Deuterium and ^{18}O-leaving-group) for substrate binding (EIE_{b1}), proton transfer to Neu5acPNP from Asp-151 (KIE_{pt1}) as well as for proton transfer from protonated product to Asp-151 (KIE_{pt2}) and unbinding of product (EIE_{ub2}). Calculated V_{max}/K_m effects are the product of EIE_{b1} and KIE_{pt1}. Results are given for both core sizes.

EFFECT	EIE_{b1} (uR to bR)	KIE_{pt1} (bR to TS1)	KIE_{pt2} (bP^+ to TS2)	EIE_{ub2} (bP to uP)	V_{max}/K_m (uR to TS1)	V_{max}/K_m (Expt)[a]
Large Core						
3-Pro-R[^2H]	0.945	1.033	0.995	1.015	0.976	1.035±0.010
3-Pro-S[^2H]	0.933	1.050	1.010	1.069	0.980	1.058±0.006
3-[$^2H]_2$	0.881	1.085	1.005	1.084	0.956	1.087±0.015
2-[^{18}O]	1.005	1.024	1.003	0.996	1.029	1.045±0.021
Small Core						
3-Pro-R[^2H]	0.945	1.033	0.995	1.015	0.976	
3-Pro-S[^2H]	0.937	1.049	1.008	1.064	0.983	
3-[$^2H]_2$	0.885	1.083	1.003	1.080	0.958	
2-[^{18}O]	1.007	1.025	1.003	0.990	1.032	

[a]Reference 2

comparing the results obtained with the small CORE. Thus, this shows that KIEs are localised effects for this system.

More concerning still, the calculated equal β-hydron effects do not correspond to the structure of TS1. The *pro-R* and *pro-S* hydron make dihedral angles about the C2—C3 bond to O_g of 22° and 137° respectively in the transition structure. This conformation of TS1 should lead to a *pro-R* effect that is greater than the *pro-S* since the overlap between the *pro-R* hydrogen and developing empty p-orbital is greater. The effect associated with either hydron should thus not be of equal value.

One possible reason for the "wrong" answer is the inadequate solvation around the reacting centre of the substrate. The O_g atom in the unbound reactant has no hydrogen bonds to it at all: no water molecules are found within 3Å. In addition to the solvation issues with uR, the enzyme-substrate complex may contain "*in vacuo*" pockets that require solvation, which could perhaps affect the substrate conformation. Overcoming these problems may involve using an approach similar to that of Cunningham *et al.*[161] An alternative method may involve resolvating the ES complex by continually rotating the system around. Such a method has been used by Richards and Mulholland in their QM/MM study of citrate synthase.[155]

Alternatively, the flattening of the boat conformation in the reactant and TS may result from interactions between the substrate and the surrounding classically described enzyme. The interactions between the arginine triad (Arg-118, Arg-292 and Arg-371) and the carboxylate group on the substrate are very strong. However, table 8.10 shows that these interactions are much closer in both enzyme-substrate complexes compared to those found in the crystal structure. These findings suggest that the QM/MM Hamiltonian may be over-estimating the interaction between the arginine triad and the carboxylate group on the substrate. For Neu5acPNP, the carboxyl oxygen atoms are held firmly in place throughout the simulation, only deviating by about 0.2Å between the bound reactant and sialyl zwitterion intermediate. To maximise the conjugative effect with O_h, the system must bring the C2 centre into the plane of the pyranose ring (figure 8.5).

Table 8.10. Distances in Ångstroms between the hydrogen bond donor nitrogen atoms of the Arginine triad and substrate carboxyl oxygen atoms. The distances refer to bR, bP and the original X-ray crystal structure. The average interaction distance for each structure is also given.

Interaction	Neu5acPNP	Neu5ac	Neu5ac (crystal)
Arg-118 NH1 - OA Substrate	2.800	2.967	3.149
Arg-118 NH2 - OA Substrate	3.644	4.163	3.696
Arg-292 NH1 - OB Substrate	2.848	2.948	2.943
Arg-292 NH2 - OB Substrate	2.826	2.857	3.190
Arg-371 NH1 - OA Substrate	2.804	2.886	3.184
Arg-371 NH2 - OB Substrate	2.827	2.811	2.835
Average Distance	2.96	3.11	3.17

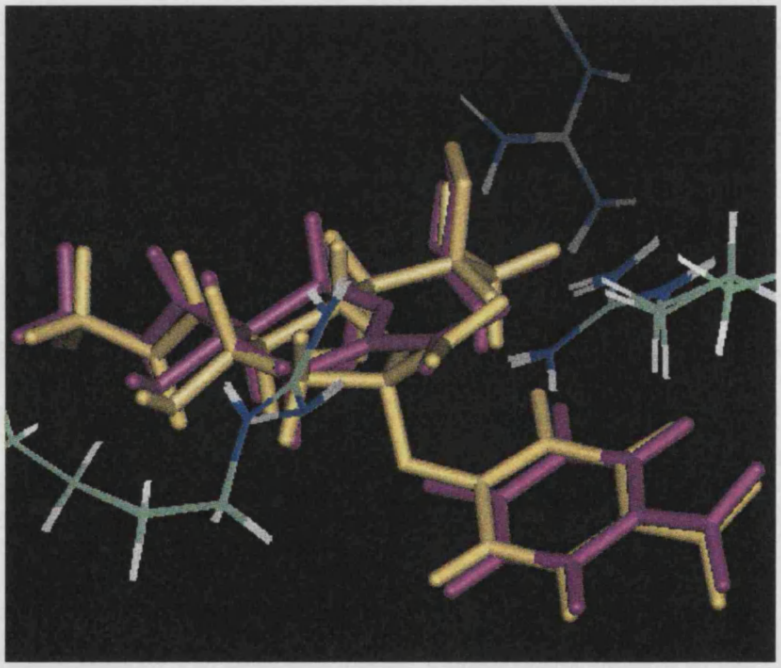


Fig 8.5. Comparison of the reactant (yellow) and sialyl zwitterion (purple) geometries for proton transfer to Neu5acPNP. The C2 centre is seen to move into the plane of the pyranose ring allowing maximum conjugation with the adjacent pyranose oxygen atom. Arg118, Arg-292 and Arg-371 are also shown.

In considering the bound product, Neu5ac, the structure is also flattened at the C2 centre, with a C6—O_h—C2—C3 dihedral of about –38°. Comparing this to the crystal structure in which the corresponding dihedral is –53° shows that while flattening of

the C2 centre is over-estimated, it is not as pronounced compared with the Neu5acPNP substrate. This perhaps suggests that the PNP leaving-group may be partly responsible for the incorrect conformation, although the C2—O_g bond in both substrates is a respectable length for a normal C—O single bond. The only protein residues to be within 3Å of the *p*-nitrophenyl leaving group are Asp-151 and Arg-292.

The KIEs presented here suggest these simulations require further development. Obviously, advancing the QM region to include more atoms around the substrate reaction centre (i.e. C6, O_h, C2, O_g, C3, the carboxylate group and β-hydrons) is a possibility. The protein residues found within 3Å of this substrate reaction centre are Arg-118, Asp119, Asp-151, Arg-292, Arg-371 and Tyr-406. However, increasing the size of the QM region to include these residues has the disadvantage of increased computational expense.

8.6. Atomic Displacements in Acid Catalysis to Neu5acPNP

Information on the flexibility of the system throughout the simulation can be obtained by comparing the atomic displacements between the TS and the local minima it connects. By only including the flexible atoms in the optimisation, efficient use of computer time in more time-consuming simulations could be achieved. Alternatively, the entire enzyme sub-unit could be modelled, allowing longer-range interactions to be included in the simulation. For the two proton transfer steps studied in section 8.3, the protonation of Neu5acPNP results in the greatest amount of structural change at the reacting centre. Therefore, this step is used to determine the flexibility of the system throughout the simulation of the PT processes.

The molecular similarity between the TS and the optimised reactant and product structures it connects was evaluated by superimposing the protein backbone atoms for all three structures. The RMS deviation for the atomic positions between the TS structure and either minimum was 0.01Å. Thus, any protein residue in the reactant or product with an overall RMS deviation of ≥0.02Å (compared to its position in the

TS) was considered "flexible" throughout the simulation. Table 8.11 lists the flexible sialidase residues in this simulation, which are all found within 7.5Å of the Neu5acPNP substrate.

Table 8.11. Sialidase residues whose overall RMS deviation of atomic positions between the TS structure and either minimum is ≥0.02Å

Residue	Reactant RMS Dev.	Product RMS Dev.	Residue	Reactant RMS Dev.	Product RMS Dev.
Arg-118	0.06	0.06	Ser-245	0.02	0.03
Glu-119	0.02	0.06	Ala-246	0.05	0.04
Gln-136	0.04	0.01	Glu-276	0.06	0.06
Val-149	0.04	0.04	Glu-277	0.03	0.06
His-150	0.07	0.04	Arg-292	0.07	0.07
Arg-152	0.25	0.08	Asn-294	0.04	0.04
Arg-156	0.02	0.02	Gln-347	0.07	0.01
Trp-178	0.03	0.03	Gly-348	0.05	0.04
Ser-179	0.04	0.01	Val-349	0.02	0.04
Ile-194	0.02	0.01	Lys-350	0.01	0.02
Asp-198	0.02	0.01	Arg-371	0.05	0.08
Ile-222	0.11	0.03	Gly-405	0.00	0.02
Arg-224	0.02	0.02	Tyr-406	0.03	0.08
Thr-242	0.01	0.02	Ile-427	0.04	0.04
Gly-244	0.01	0.03	Lys-431	0.05	0.02

Solvent molecules are likely to be more flexible than the enzyme. Indeed, when comparing the individual atomic positions of all the waters, the "flexible" solvent molecules were found within 15Å of the Neu5acPNP substrate. Thus, a system that incorporates flexible protein residues and solvent molecules within 7.5Å and 15Å of the substrate, respectively, will have about 1100 optimisable atoms (*cf.* 400 in Jourand's work). It should be noted that these findings only apply to the proton transfer step studied here; other mechanistic investigations involving alternative steps may require different parts of the enzyme to be flexible in the simulation.

Using a large, flexible system (as in the simulations presented in this chapter) allows insight into which atoms in the system require optimisation.

8.7. Conclusions

QM/MM hybrid modelling has been used by several groups to investigate a variety of enzyme mechanisms, giving valuable insight into possible mechanistic pathways in such systems. The work presented in this chapter uses the QM/MM hybrid modelling technique to look at the mechanism of influenza-A sialidase.

From several experimental and theoretical investigations on the action of sialidase undertaken, a mechanism has been postulated for a retaining pathway. This involves general acid catalysis by Asp-151 resulting in the formation of an enzyme bound sialyl zwitterion, followed with its destruction by nucleophilic attack of water at the positively charged centre.

The TS structures for proton transfers from Asp-151 to either the Neu5acPNP or Neu5ac substrates involve late proton donation along with early cleavage of the glycoside bond. It has been shown with the AM1/CHARMM potential that increasingly facile leaving groups cleave away from the sialyl moiety in concert with the proton donation from Asp-151. Using these findings, a lactose moiety is suggested to require protonation before its departure. It was also found that the TS for proton transfer to Neu5acPNP was stabilised by the sialidase enzyme, along with the resulting sialyl zwitterion.

The calculated KIEs for the effect on V_{max}/K_m are found to be in disagreement with the experimental results obtained by Guo *et al.*[2] This is in part due to the wrong conformation of the TS structure, which is related to a reactant in a "near half-chair" conformation rather than a $B_{2,5}$ conformation. The unreliability of the QM/MM procedure here has been attributed to a) inadequate solvation of the uR and bR structures; b) over-estimation in the QM/MM interactions between the arginine triad and the substrate carboxylate group; and c) suspected poor performance of the AM1 Hamiltonian in modelling the chemical changes investigated within this work. For example, AM1 was found to over-estimate the energy change associated with the

cleavage of an ether bond (compared to the MP2/6-31+G(d) level of theory) by *ca.* 80 kJ mol^{-1}. This may be alleviated in part by the advanced proton transfer in the TS for general acid catalysis found using this level of theory.

Finally, some improvements to the simulations have been suggested including many-step solvation techniques and the use of a larger QM system around the substrate reaction centre. The latter is likely to dramatically increase the computational expense in modelling the sialidase system. However, it has been shown for the proton transfer to Neu5acPNP that "major displacements" involve protein residues and solvent molecules within 7.5Å and 15Å of the substrate respectively. Using this information to fix "non-flexible" atoms in the system would alleviate some of the computational expense associated with having a large QM region.

The modelling problems presented in this chapter would not have been apparent without experimental results. QM/MM hybrid modelling can readily provide a wrong answer as demonstrated with the influenza-A sialidase work presented here. The interplay between theory experiment is therefore a very important part of the QM/MM hybrid modelling process.

9. CONCLUSIONS AND POSSIBILITIES FOR FURTHER STUDY

The structure and stability of α-lactones has been considered extensively in this thesis. It is evident from published data on the small isolatable molecules that substituents at the α-centre do affect the stability (and ring strain) in such systems. Electron-donating substituents are able to stabilise the cationic centre and promote the formation of a zwitterionic species as opposed to a strained α-lactone. Electron-withdrawing species, on the other hand, will promote the formation of an α-lactone; in the extreme case, nucleophilic attack on α-lactones incorporating such groups will preferentially occur at the carbonyl centre.

The way by which substituents at the α-centre alter the structure and stability of the three-membered ring may be (indirectly) obtained through the position of the carbonyl IR absorption. More stable α-lactones will have their carbonyl stretch shifted to higher frequencies, whereas less stable species will have such absorptions closer to the carbonyl stretches associated with carboxylic acids, esters, etc. Theoretical studies are able to calculate vibrational frequencies and this provides a convenient way to (indirectly) compare theoretical structure-activity information with available experimental data.

Theory is able to give information on how α-substituents increase or decrease the ring strain within an α-lactone. In the case of hydroxyoxiranone, for example, the hydroxyl group is found to reduce the ring strain of the three membered ring through n_{π}—p_{π} interactions, resulting in the loosening of the C_{α}—O_{n} bond. The hydroxyl group may therefore reduce the carbonyl stretching frequency as compared to the unsubstituted oxiranone molecule. The interplay between theory and experiment is important, but the lack of experimental results often precludes such a comparison to be made. Test simulations (to determine the accuracy of the calculated carbonyl absorption associated with many experimentally determined α-lactones) will not give definitive prove that the calculated structure of the system is correct. However,

such simulations do provide the opportunity to calibrate the vibrations in the models being used.

Using a carbonyl-absorption calibrated model, it may be possible to undertake a theoretical structure-activity relationship (SAR) study to investigate substituent effects upon ring strain in α-lactones and then extend this study to investigate solvation effects. Such simulations might be achieved using continuum models (that allow optimisation) using different dielectric constants. However, α-lactones in an aqueous environment could also be modelled using a hybrid QM/MM technique. Two possibilities exist here: firstly, to use a semi-empirical QM description of the α-lactone within a TIP3P environment, or alternatively, to use an *ab initio* description of the solute. The latter is possible because α-lactones usually consist of a small number of atoms (see chapter 4). The *ab initio*/TIP3P description could be achieved by using Turner's GRACE program.[27]

When considering intermediates in non-enzymic sialoside hydrolysis, appropriate conformations of the small models were used to represent the larger sialyl system. Using the smaller model allowed higher levels of theory to be employed on a realistic CPU time scale. Three intermediates were considered in this thesis: an α-lactone (represented by hydroxyoxiranone), a zwitterion and carbene. Each of the small models included a hydroxy substituent on the α-position (section 5.1). This group played an essential role in stabilising the reacting centre, as already discussed with hydroxyoxiranone. However, the stabilisation afforded to the zwitterion and carbene intermediates by the hydroxyl substituent was much more effective. This is presumably because an empty p-orbital already exists at the reaction site allowing for increased n_π—p_π interactions as compared with the partially filled p-orbital of hydroxyoxiranone. The pyranose ring in the full sialyl system therefore plays an important role in stabilising electron deficient intermediates. In addition to this effect, hyperconjugation and conformational aspects may also stabilise the system, as evidenced from the exothermic energy change when restraining the smaller model into a ring.

For the smaller model, the α-lactone is the most stable species of the three intermediates *in vacuo*, with the charge-separated zwitterion being the highest in energy. In solution, the zwitterion is greatly stabilised by the surrounding solvent. However, using the small hydroxyl models within the IPCM continuum model, affords an α-lactone that is still lower in energy than the zwitterion in the aqueous environment. Caution must therefore be taken when using small and/or continuum models.

A great deal of evidence has been provided in this thesis to suggest that the α-lactone is actually higher in energy than a zwitterion intermediate in an aqueous environment. The large *ab initio* models that incorporate the pyranose ring show this in the presence of the continuum solvent; semi-empirical AM1 descriptions with the COSMO continuum model also agree with this conclusion. Furthermore, using the AM1/TIP3P description, all ten α-lactone/zwitterion pairs showed the zwitterion to be more stable than the α-lactone in an aqueous environment.

The energy change in going from a sialyl α-lactone intermediate to the corresponding zwitterion using the AM1/TIP3P hybrid method was –60.1 ± 32.6 kJ mol^{-1}. The large error bar was simply due to the many non-averaged solvent-solvent interactions present within the system, as pointed out by Kollman.[101] The α-lactone-zwitterion energy difference calculated by Horenstein would be contained within these error bars,[76] although such calculations used an *ab initio* description of the (cut-down) solute model and a continuum description of the solvent.[76]

It was shown in chapter 6 that there existed a "hysteresis effect": the energy difference between the two different intermediates depended on the direction taken along the profile for α-lactone ring opening (or ring closure). The average value obtained in going from α-lactone to zwitterion was –30.5 kJ mol^{-1} whereas in going from zwitterion to α-lactone the calculated energy change was –89.6 kJ mol^{-1}. The value associated with the hysteresis effect was thus as much as the overall average energy difference (*ca.* 60 kJ mol^{-1}). This difference has been mostly attributed to the different *immediate* solvent environments generated around the two different solute intermediates. With such a large value for the hysteresis effect, it is difficult to

conclude the actual value for the difference between the α-lactone and the zwitterion intermediates.

A possible way to minimise the hysteresis effect is to have a more consistent *immediate* solvent structure for all ten runs. Rather than initially solvating both reactants and products, it may be possible to start from a TS structure for the α-lactone ring opening process. This, of course, assumes that the TS for a specific chemical change belongs to a family of similar transition structures as suggested by Moliner *et al.*[103] The problem to be solved is where to start? Perhaps the simplest way is to provide a guess TS structure using the AM1/COSMO method (*ts0*). This could be solvated using the *immediate* solvent sphere (of 10Å radius?) with subsequent generation of several solvent structures around *ts0*. The TS starting structures can then be produced by adding the *bulk* solvent sphere (as described for the α-lactone/zwitterion starting points in chapter 6). The resulting structures would then have to be subsequently refined to make sure they lie on a first order saddle point. IRC calculations could then be used to generate both α-lactone and zwitterion structures connected to the TS (followed by optimisation of the system). This suggested process is schematically shown in figure 9.1. However, although this method could reduce the hysteresis effect, it may also reduce the energy barriers associated with both intermediates since the TS is preferentially stabilised by the solvent over the two minima structures.

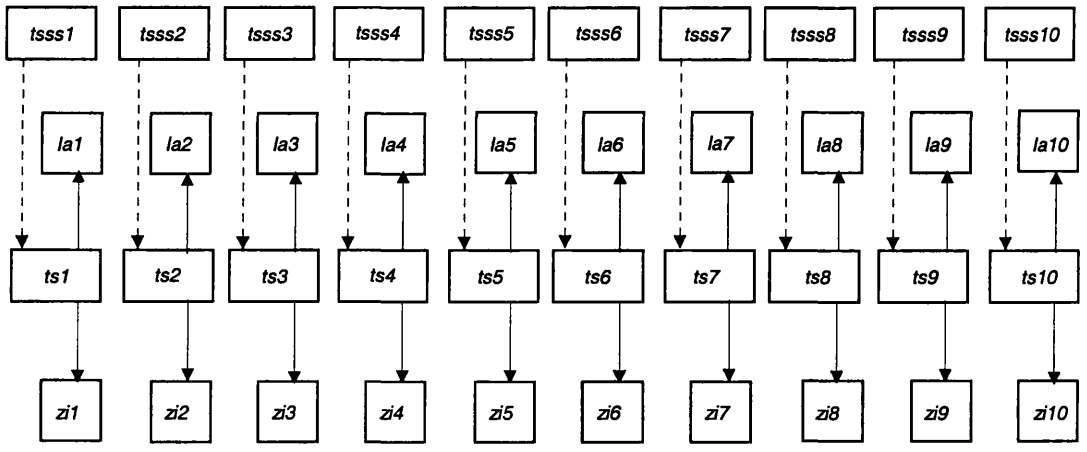


Fig 9.1. A possibility for the future. Start from a guess TS structure (perhaps obtained with the AM1/COSMO method) and generate several transition-state starting structures (*tsss*). Refine each starting point in order to make sure the structure is on the first-order saddle point (dotted arrow). From each TS, move to the local reactant and product wells using an IRC method (solid arrow), followed by optimisation of the resulting intermediate to ensure the minimum well has been reached.

One of the other problems associated with this work is the restricted conformation sampling done on the system: Monte-Carlo (MC) techniques, for example, usually generate millions of configurations. The work presented in this thesis only uses ten starting structures. With more computation time, it would be possible to look at more starting points, but the analysis of such large simulations rapidly becomes unwieldy. Considering that ten runs used some 400 hours of CPU time on the DEC 8400 at the Rutherford-Appleton laboratory, investigating a million runs in this manner would be computationally intractable. MC techniques may be able to provide 50 low energy configurations of the solvent environment as the basis for more simulations. Alternatively, the sampling of conformational space could be done using simulated annealing procedures that cool the system from elevated temperatures (*ca.* 1000K and higher). However, the different nature of the two intermediates (the α-lactone is a strained species with no formally separated charges, whereas the zwitterion is a charge-separated system) may still lead to a hysteresis effect.

The hysteresis problem may also lie with how the simulation is conducted. IRC calculations on the enzyme system, for example, only result in local changes around the reaction centre while the rest of the system remains essentially fixed. As shown in section 8.8, only a third of the system shows an RMS change of ≥0.02Å in going from local minima to TS. Admittedly, the enzyme environment is more structurally restrained than a solvent-only environment. However, the limited change of the IRC calculation may not allow full solvent relaxation around the "drastically changing" solute moiety. The option of moving from one intermediate to another using a very large number of optimisation cycles (*ca* 4000 cycles per 0.05Å elongation/shortening of the C_α—O_n bond) could be investigated. This may prove valuable in determining whether simple optimisation procedures alone can allow the necessary solvent relaxation for modelling chemical reactivity.

The "multiple starting structure approach" could perhaps be used to also generate ten starting structures when considering the influenza sialidase mechanism. An initial enzyme-substrate TS structure could be first determined using simulations similar to those presented within this thesis. Because the enzyme system is larger and more rigid than in the case of the solvent-only environment, the amount of simulated

annealing required will also be larger (to ensure adequate sampling of conformation space). One possible approach to generating the starting structures could involve fixing the TS CORE and subjecting the ENVIRONMENT to a period of MD. This simulation may require heating the system to 310K (37°C, the temperature at which the experimental KIEs were determined) over, say, 20ps and then equilibrating it (for *ca.* 80ps?). After successful equilibration, a production run of about 150ps could be run. Starting structures could then be extracted from the trajectory (separated by 15ps intervals) with each structure being subsequently cooled to absolute zero. The cooling time to CPU time ratio would need to be maximised by doing a few test runs before embarking on the main simulation. After the cooling period, each structure could then be optimised around the fixed CORE until a specific gradient tolerance (*ca.* 0.02 kJ mol^{-1}) has been achieved. Subsequent refinement of the TS structure would then ensure the starting structure is located at a first-order saddle point.

However, as discussed in chapter 8, there presently exist many problems with enzyme QM/MM simulations; these need to be ironed out before any new simulations are attempted. Many tests are now required to try to understand where these problems lie.

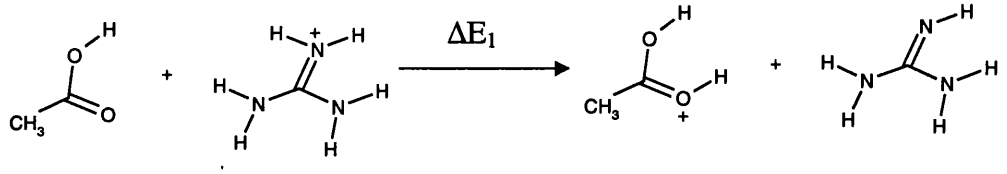
One problem is solvation. The unbound substrate and ES complex both need to be adequately solvated. Two possibilities of achieving this exist here. The first involves a method of adding solvent molecules around the "solute" in many stages. Solvent molecules could be added to the bare "solute" followed by deletion of the overlapping water molecules and optimisation of the system. The resulting structure could then be rotated by a number of degrees (*ca.* 60°?) before adding a further set of solvent molecules. This stage would require the deletion of new overlapping waters, but the retention of water molecules filling "*in vacuo* cavities" present within the system. The structure would then be subjected to further optimisation. This process could be repeated until no further water molecules are added to the system. A related method was used in the citrate synthase work of Mulholland and Richards.[155] The second approach subjects the system to a period of simulated annealing before adding a second set of water molecules; this process could be repeated until no further solvent molecules are added to the system. A similar process was used by Cunningham *et al.* in their QM/MM study of the MDH mechanism.[161]

Another potential problem is the interaction between the QM and MM regions. Are electrostatic interactions over-estimated? In the sialidase system, there is evidence to suggest that the interaction between the QM described carboxylate group of the substrate is interacting too strongly with the adjacent MM described arginine triad. To confirm whether this is the case, a potential could be mapped out involving the gradual removal of a model carboxylate group (in discrete steps) from a model arginine triad. This could be compared to similar gas phase profiles generated at various different *ab initio* levels of theory to determine whether the QM/MM partition describes the interaction effectively. The models employed for such simulations might, for example, be a methanoate anion for the substrate carboxylate group and a guanidinium ion for each arginine residue.

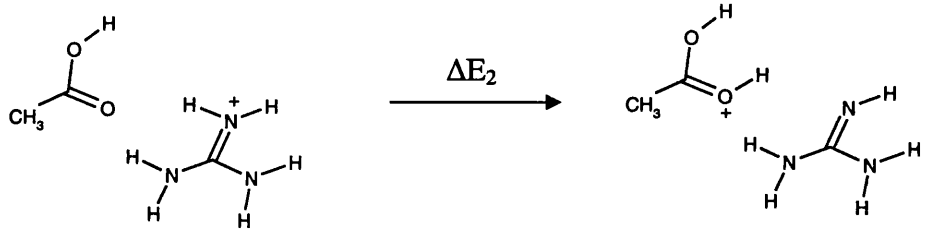
The QM system could be extended to push back the QM/MM partition from the reacting centre. Such a centre could be defined as the C6, O_h, C2, C3, the carboxylate group and β-hydrogen atoms of the substrate. The side-chains belonging to the enzyme residues within, say, 3Å of the reaction centre could be incorporated into the QM system. This would probably result in a region containing the substrate, Asp-151, the arginine triad, and Tyr-406. Including the side chains of these residues would result in 125 QM atoms. This would be very computationally intensive; however, it was shown in chapter 8 that the optimisation process could be restricted to include only 1100 atoms in the system without too much loss in flexibility.

However, is AM1 able to describe adjacent residues in the active site correctly? Jourand provided evidence that it could not: simulations incorporating both Arg-292 and adjacent Glu-277 side-chains into the QM region resulted in a spontaneous proton transfer from the basic residue to Glu-277.[163] When investigating the formation of a covalent intermediate, in which the tyrosine hydrogen is donated to the other oxygen of the glutamate, Jourand subsequently obtained a doubly protonated residue.[163] In considering a related model acid-base pair, Lapthorn showed how the energetics associated with the proton transfer altered when considering the acid-base pair over the infinitely separated species (table 9.1).[175] The *in vacuo* energy change associated with eqn. 9.1 for the infinitely separated species

was calculated to be about 230 kJ mol^{-1}.[175] Interestingly, the energy change associated with the proton transfer in the intimate acid-base pair was found to be about 0 kJ mol^{-1} (in the gas phase).[175] In comparison, the corresponding change using the PM3 semi-empirical method gave changes of 190 kJ mol^{-1} and 100 kJ mol^{-1} for the separated species and acid-base pairs respectively.[175] Bash and co-workers have reparameterised AM1 to accurately reproduce *ab initio* structures and experimentally determined energies for a particular chemical process.[161,99,100] Incorporating more enzyme-substrate residues into the QM system for the influenza sialidase work may also necessitate the reparameterisation of the AM1 method.



(9.1)



(9.2)

Table 9.1. The *in vacuo* energy difference (kJ mol^{-1}) associated with a transferring proton between a guanidinium ion and acetic acid. Both species are either described as being infinitely separated (ΔE_1) or as being an acid-base pair (ΔE_2) using the AM1 or PM3 Hamiltonian.

Energy Change	AM1	PM3
ΔE_1	228.8	191.6
ΔE_2	0.0	100.8

Another issue to be considered is where to cut the enzyme: have the simulations here incorporated too little of the enzyme environment? Does the cut need to be done at a larger distance (*ca.* 20Å radius) from the active site? Do all enzyme atoms need to be included (but not necessarily optimised)? The exclusion of some longer-range

interactions could conceivably contribute to the incorrect conformation of the substrate.

QM/MM modelling of enzyme mechanisms is obviously plagued with problems. If no experimental results are available, these problems are not apparent. For effective and reliable QM/MM modelling of enzyme mechanisms, therefore, it is *essential* to have the interplay between theory and experiment. When such data exists, QM/MM methods have been shown to work well. Alex and Finn have investigated relative binding energies of thermolysin inhibitors using a QM/MM potential (not involving optimisation).[176] The inhibitor was described with a semi-empirical Hamiltonian, while the thermolysin protein was treated classically "as a set of point charges and van der Waals spheres." The experimental interplay involves the relative difference between pK_i values for several different inhibitors. Such simulations using a QM/MM methodology were found to perform well and they provide a promising start for the use of QM/MM techniques in the pharmaceutical industry.

Despite its difficulties, QM/MM simulations of enzyme systems have the ability to provide valuable and reliable information when treated with care. As computers increase in speed, more complex QM methods and MM force fields will be available to provide a more accurate description of the system. For QM/MM modelling in general, the future can only be bright.

10. REFERENCES

1. Ashwell, M., Guo, X., Sinnott, M.L. (1992). Pathways for the hydrolysis of glycosides of N-acetyl neuraminic acid. *J. Am. Chem. Soc.*, **114**, pp. 10158-10166

2. Guo, X., Laver, W.G., Vimr, E., Sinnott, M.L. (1994). Catalysis by two sialidases with the same protein fold but different stereochemical courses: A mechanistic comparison of the enzymes from influenza A virus and *salmonella typhimurium*. *J. Am. Chem. Soc.*, **116**, pp. 5572-5578

3. Leach, A.R. (1996). Molecular modelling: principles and applications. Longman, Harlow

4. Maskill, H. (1985). The physical basis of organic chemistry. Oxford University Press, Oxford

5. Hinchliffe, A. (1996). Modelling molecular structures. Wiley, Chichester

6. Atkins, P.W., Friedman, R.S. (1997). Molecular Quantum Mechanics (3rd Ed.). Oxford University Press, Oxford

7. Foresman, J.B., Frisch, Æ. (1995). Exploring chemistry with electronic structure methods (2nd Ed). Gaussian, inc., Pittsburgh

8. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Gill, P.M.W., Johnson, B.G., Robb, M.A., Cheeseman, J.R., Keith, T., Petersson, G.A., Montgomery, J.A., Raghavachari, K., Al-Laham, M.A., Zakrzewski, V.G., Ortiz, J.V., Foresman, J.B., Peng, C.Y., Ayala, P.Y., Chen, W., Wong, M.W., Andres, J.L., Repogle, E.S., Gomperts, R., Martin, R.L., Fox, D.J., Binkley, J.S., Defrees, D.J., Baker, J., Stewart, J.J.P., Head-Gordon, M., Gonzalez, C., Pople, J.A., Gaussian 94, Revision B3 and C3, Gaussian, Inc., Pittsburgh, PA, 1995.

9. Pople, J.A., Head-Gordon, M., Raghavachari, K. (1987). Quadratic configuration interaction. A general technique for determining electron correlation energies. *J. Chem. Phys.*, **87**, pp. 5968-5975

10. Dewer, M.J.S., Zoebisch, E.G., Healy, E.F., Stewart, J.J.P. (1985). AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, **107**, pp. 3902-3909

11. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.*, **4**, pp. 187-217

12. MacKerell, Jr., A.D., Wiórkiewicz-Kuczera, J., Karplus, M. (1995). An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.*, **117**, pp. 11946-11975

13. MacKerell, Jr., A.D., Bashford, D., Bellott, M., Dunbrack, Jr., R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, III, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Wantanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., Karplus, M. (1998). All atom empirical potential for molecular modelling and dynamics studies of proteins. *J. Phys. Chem.*, **102**, pp. 3586-3616

14. Field, M.J., Bash, P.A., Karplus, M. (1990). A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.*, **11**, pp. 700-733

15. Brooks, III, C.L., Karplus, M. (1983). Deformable stochastic boundaries in molecular dynamics. *J. Chem. Phys.*, **79**, pp. 6312-6325

16. Foresman, J.B., Keith, T.A., Wiberg, K.B., Snoonian, J., Frisch, M.J. (1996). Solvent Effects. 5. Influence of cavity shape, truncation of electrostatics, and electron correlation on *ab initio* reaction field calculations. *J. Phys. Chem.*, **100**, pp. 16098-16104

17. Klamt, A., Schüürmann, G. (1993). COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2.*, pp. 799-805

18. Fukui, K. (1981). The path of chemical reactions – the IRC approach. *Acc. Chem. Res.*, **14**, pp. 363-368

19. Cornish-Bowden, A. (1995). Fundamentals of Enzyme Kinetics (Rev. Ed.). Portland Press, London.

20. Michaelis, L., Menten, M.L. (1913). *Biochem. Z.*, **49**, pp. 333-369

21. Williams, I.H. (1992). CAMISO, University of Bath, Bath

22. Frisch, M.J., Trucks, G.W., Head-Gordon, M., Gill, P.M.W., Wong, M.W., Foresman, J.B., Johnson, B.G., Schlegel, H.B., Robb, M.A., Repogle, E.S., Gomperts, R., Andres, J.L., Raghavachari, K., Binkley, J.S., Gonzalez, C., Martin, R.L., Fox, D.J., Defrees, D.J., Baker, J., Stewart, J.J.P., and Pople, J.A. Gaussian 92, Revision C.4, Gaussian, Inc., Pittsburgh, PA., 1992.

23. MOPAC93, Stewart, J.J.P., and Fujitsu Limited, Tokyo, Japan, 1993.

24. Varghese, J.N., McKimm-Breschkin, J.L., Caldwell, J.B., Kortt, A.A., Colman, P.M. (1992). The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins.*, **14**, pp. 327-332

25. QUANTA, Molecular Simulations Inc., 200 Fifth Avenue, Waltham, MA 02154, USA

26. Hart, J.C., Burton, N.A., Hillier, I.H., Harrison, M.J., Jewsbury, P. (1997). Prediction of transition state structure in protein tyrosine phosphatase catalysis using a hybrid QM/MM potential. *Chem. Commun.*, pp. 1431-1432

27. Turner, A.J. (1997). Combined quantum/classical modelling of chemical reactions in enzymes and solution. *PhD thesis*, University of Bath.

28. Williams, I.H. (1983). Force-constant computations in Cartesian coordinates. Elimination of translational and rotational contributions. *J. Mol. Struct. (THEOCHEM)*, **94**, pp. 275-284

29. Williams, I.H. (1982). On the representation of force fields for chemically reacting systems. *Chem. Phys. Lett.*, **88**, pp. 462-466

30. Taylor, N.R., von Itzstein, M. (1994). Molecular modeling studies on ligand binding to sialidase from influenza virus and the mechanism of catalysis. *J. Med. Chem.*, **37**, pp. 616-624

31. Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M., Hagler, A.T. (1988). Structure and energetics of ligand binding to proteins: *Escherichia coli* dihydrofolate reductase-trimethoprim, a drug-receptor system. *Proteins*, **4**, pp. 31-47

32. von Baeyer, A. (1885). Ueber polyacetylenverbindungen. *Chem. Ber.*, **18**, pp. 2269-2281

33. Rodriquez, C.F., Williams, I.H. (1997). Ring strain energy and enthalpy of formation of oxiranone: an *ab initio* theoretical determination. *J. Chem. Soc. Perkin Trans. 2.*, pp. 953-957

34. Milligan, D.E., Jacox, M.E. (1962). Infrared study of the reaction of CH_2 with CO_2 in the solid state. *J. Chem. Phys.*, **36**, pp. 2911-2917

35. Chapman, O.L., Wojtkowski, P.W., Adam, W., Rodriquez, O., Rucktäschel, R. (1972). Synthesis and chemistry of α-lactones. *J. Am. Chem. Soc.*, **94**, pp. 1365-1367

36. Sander, W.W. (1989). Reaction of diphenylmethylene with carbon dioxide: matrix isolation of diphenyloxiranone. *J. Org. Chem.*, **54**, pp. 4265-4267.

37. Adam, W., Liu, J.-C., Rodriguez, O. (1973). Bis(trifluoromethyl)acetolactone, a stable α-lactone. *J. Org. Chem.*, **38**, pp. 2269-2270

38. Coe, P.L., Sellars, A., Tatlow, J.C., Whittaker, G., Fielding, H.C. (1982). An isolatable α-lactone made by direct epoxidation. *J. Chem. Soc. Chem. Commun.*, pp. 362-363

39. Wierlacher, S., Sander, W., Liu, M.T.H. (1992). Carboxylation of carbenes in low-temperature matrices. *J. Org. Chem.*, **57**, pp. 1051-1053

40. Adam, W., Rucktäschel,R. (1971). Cyclic peroxides, V. An α-lactone intermediate via photodecarboxylation of a monomeric malonyl peroxide. *J. Am. Chem. Soc.*, **93**, pp. 557-559

41. Wheland, R., Bartlett, P.D. (1970). α-Lactones from diphenylketene and di-*tert*-butylketene. *J. Am. Chem. Soc.*, **92**, pp. 6057-6058

42. Coe, P.L., Owen, I.R., Sellars, A. (1989). Reactions of tetrafluoroethane oligomers. Part 9. Some reactions of perfluoro-(1-ethyl-1-methylpropyl)(s-butyl)ethanolide (an α-lactone). *J. Chem. Soc. Perkin Trans. 1.*, pp. 1097-1103

43. Evans, B.W., Fenn, D.J., Tighe, B.J. (1970). Studies of the reactions of the anhydrosulphites of α-Hydroxy-carboxylic acids. Part V. Thermal

decomposition of benzilic and mandelic acid anhydrosulphites. *J. Chem. Soc. (B).*, pp. 1049-1052

44. Markgraf, J.H. (1998). Infrared carbonyl frequencies of heterocyclic lactones. *Heterocycles.*, **47**, pp. 559-624

45. Bellamy L.J. (1974). The infrared spectra of complex molecules (3rd Ed.), Chapman-Hall, London, pp. 211-214.

46. Winstein, S., Lucas, H.J. (1939). Retention of configuration in the reaction of the 3-bromo-2-butanols with hydrogen bromide. *J. Am. Chem. Soc.*, **61**, pp. 1576-1581

47. Cowdrey, W.A., Hughes, E.D., Ingold, C.K. (1937). Reaction kinetics and the Walden inversion. Part III. Homogeneous hydrolysis and alcoholysis of α-bromopropionic acid, its ester and anion. *J. Chem. Soc.*, pp. 1208-1225

48. Cowdrey, W.A., Hughes, E.D., Ingold, C.K., Masterman, S., Scott, A.D. (1937). Reaction kinetics and the Walden inversion. Part VI. Relation of steric orientation to mechanism in substitutions involving halogen atoms and simple or substituted hydroxyl groups. *J. Chem. Soc.*, pp. 1252-1271

49. Ballard, D.G.H., Tighe, B.J. (1967). Studies of the reactions of the anhydrosulphites of α-Hydroxy-carboxylic acids. Part I. Polymerisation of anhydrosulphite of α-Hydroxyisobutyric acid. *J. Chem. Soc (B).*, pp. 702-709

50. Cameron, T.B., El-Kabbani, F.M., Pinnick, H.W. (1981). Flash vacuum thermolysis of 1,3-dioxolan-4-ones. *J. Am. Chem. Soc.*, **103**, pp. 5414-5417

51. Adam, W., Cadiz, C., Mazenod, F. (1981). Vacuum flash pyrolysis (VFP) of malonyl peroxides: Decarboxylation verses decarbonylation of the intermediary α-lactones. *Tet. Lett.*, **22**, pp. 1203-1206

52. Gortler, L.B., Saltzman, M.D. (1966). Peroxy esters of dimethylmalonic acid. α-Lactone intermediates in free-radical reactions. *J. Org. Chem.*, **31**, pp. 3821-3823

53. Walling, C., Savas, E.S. (1960). The reaction of phenyl radicals with carbon monoxide; some observations on the induced decomposition of benzoyl peroxide. *J. Am. Chem. Soc.*, **82**, pp. 1738-1744

54. Bartlett, P.D., Gortler, L.B. (1963). Peresters. XI. Di-*tert*-butylperoxy diphenylmalonate. α-Lactone intermediates in free radical reactions. *J. Am. Chem. Soc.*, **85**, pp. 1864-1869

55. Bartlett, P.D., McCluney, R.E. (1983). Autoxidatation of diphenylketene. 1. Conditions and products. *J. Org. Chem.*, **48**, pp. 4165-4168

56. Bartlett, P.D., McCluney, R.E. (1984). Correction. *J. Org. Chem.*, **49**, pp. 1472

57. Crandall, J.K., Sojka, S.A., Komin, J.B. (1974). Reactions of ketenes with peracids and ozone. *J. Org. Chem.*, **39**, pp. 2172-2175

58. Moriarty, R.M., Gupta, S.C., Hu, H., Berenschot, D.R., White, K.B. (1981). Oxygen-atom transfer from iodosobenzene to ketenes, α-keto carboxylic acids, and ketones. *J. Am. Chem. Soc.*, **103**, pp. 686-688

59. Niwayama, S., Noguchi, H., Ohno, M., Kobayashi, S. (1993). On the mechanism of the Meinwald rearrangement of electron deficient systems. *Tet. lett.*, **34**, pp 665-668

60. Niwayama, S., Kobayashi, S., Ohno,M. (1994). A novel chemicoenzymatic rearrangement by asymmetric hydrolysis with Pig Liver Esterase. *J. Am. Chem. Soc.*, **116**, 3290-3295

61. Tomioka, H., Hirai, K., Tabayashi, K., Murata, S., Izawa, Y., Inagaki, S., Okajima, T. (1990). Neighbouring group participation in carbene chemistry. Effect of neighbouring carboxylate group on carbene reactivities. *J. Am. Chem. Soc.*, **112**, 7692-7702

62. Pople, J.A., Luke, B.T., Frisch, M.J., Binkley, J.S. (1985). Theoretical thermochemistry. 1. Heats of formation of neutral AH_n molecules (A=Li to Cl). *J. Phys. Chem.*, **89**, pp. 2198-2203

63. Kolos, W., Wolniewicz, L. (1968). Improved ground state energy of the hydrogen molecule. *J. Chem. Phys.*, **49**, pp. 404

64. Lias, S.G., Bartmess, J.E., Liebman, J.F., Levin, R.D., Mallard, W.G. (1988). *J. Phys. Chem. Ref. Data, supplement 1.*

65. Report of the CODATA task group. (1978). *J. Chem. Thermodyn.*, **10**, 903

66. Hartwell, E.J., Richards, R. E., Thompson, H.W. (1948). The vibration frequency of the carbonyl linkage. *J. Chem. Soc.*, pp. 1436-1441

67. Barrow, G.M. (1953). Conjugation and intensity of the infrared carbonyl band. *J. Chem. Phys.*, **21**, pp. 2008-2011

68. Sinnott, M.L., Jencks W.P. (1980). Solvolysis of D-glucopyranosyl derivatives in mixtures of ethanol and 2,2,2-trifluoroethanol. *J. Am. Chem. Soc.*, **102**, pp. 2026-2032

69. Jencks, W.P. (1981). How does a reaction choose its mechanism? *Chem. Soc. Rev.*, **10**, pp. 345-375

70. Jencks, W.P. (1980). When is an intermediate not an intermediate? Enforced mechanisms of general acid-base catalyzed, carbocation, carbanion and ligand exchange reactions. *Acc. Chem. Res.*, **13**, pp. 161-169

71. Sinnott, M.L. (1990). Catalytic mechanisms of enzymic glycosyl transfer. *Chem. Rev.*, **90**, pp. 1171-1202

72. Amyes, T.L., Jencks, W.P. (1989). Lifetimes of oxocarbenium ions in aqueous solution from common ion inhibition of the solvolysis of α-azido ethers by added azide ion. *J. Am. Chem. Soc.* **111**, pp. 7888-7900

73. Banait, N.S., Jencks, W.P. (1991). Reaction of anionic nucleophiles with α-D-glucopyranosyl fluoride in aqueous solution through a concerted, A_ND_N (S_N2) mechanism. *J. Am. Chem. Soc.*, **113**, pp. 7951-7958

74. Horenstein, B.A., Bruner, M. (1998). The N-acetyl neuraminyl oxecarbenium ion is an intermediate in the presence of anionic nucleophiles. *J. Am. Chem. Soc.*, **120**, pp. 1357-1362

75. Horenstein, B.A., Bruner, M. (1996). Acid-catalyzed solvolysis of CMP-N-acetyl neuraminate: evidence for a sialyl cation with a finite lifetime. *J. Am. Chem. Soc.*, **118**, pp. 10371-10379

76. Horenstein, B.A. (1997). Quantum mechanical analysis of an α-carboxylate-substituted oxocarbenium ion. Isotope effects for the formation of the sialyl cation and the origin of an unusually large secondary ^{14}C isotope effect. *J. Am. Chem. Soc.*, **119**, pp. 1101- 1107

77. Bruner, M., Horenstein, B.A. (1998). Isotope trapping and kinetic isotope effect studies of rat liver α-(2→6)-sialyltransferase. *Biochemistry*, **37**, pp. 289-297

78. Norris, K.E., Bacskay, G.B., Gready, J.E. (1993). Theoretical study of "protonated pyruvate": a methylhydroxycarbene-carbon dioxide complex-implications for the decarboxylation of pyruvic acid. *J. Comput. Chem.*, **14**, pp. 699-714.

79. Briner, K., Vasella, A. (1989). Glycosylidene carbenes. A new approach to glycoside synthesis. Part 1. Preparation of glycosylidene-derived diaziridines and diazirines. *Helv. Chim. Acta.*, **72**, pp. 1371-1382.

80. Vasella, A., Briner, K., Soundararajan, N., Platz, M.S. (1991). A laser flash photolysis derived study of a glycosylidene carbene. *J. Org. Chem.*, **56**, pp. 4741-4744

81. Vasella, A., Waldraff, C.A.A. (1991). Glycosylidene carbenes. Part 3. Synthesis of spirocyclopropanes. *Helv. Chim. Acta.*, **74**, pp. 585-593

82. Bock, C.W., Redington, R.L. (1988). Isomerization and unimolecular dissociation channels of the glyoxylic acid monomer. *J. Phys. Chem.*, **92**, pp. 1178-1187

83. Altmann, J.A., Csizmadia, I.G., Yates, K., Yates, P. (1977). An *ab initio* study of the rearrangement of carbonyl compounds to oxacarbenes. *J. Chem. Phys.*, **66**, pp. 298-302

84. Harding, L.B., Schlegel, H.B., Krishnan, R., Pople, J.A. (1980). Møller-Plesset study of the H_4CO potential energy surface. *J. Phys. Chem.*, **84**, pp. 3394-3401

85. Pople, J.A., Raghavachari, K., Frisch, M.J., Binkley, J.S., Schleyer, P.v.R. (1983). Comprehensive theoretical study of isomers and rearrangement barriers of even-electron polyatomic molecules H_mABH_n (A,B=C,N,O, and F). *J. Am. Chem. Soc.*, **105**, pp. 6389-6398

86. Antolovic, D., Shiner, V.J., Davidson, E.R. (1988). Theoretical study of α-lactone, acetoxy diradical, and the gas-phase dissociation of the chloroacetate anion. *J. Am. Chem. Soc.*, **110**, pp. 1375-1381

87. Maggiora, G.M., Williams, I.H. (1982). Intermolecular interaction energies from minimal-basis SCF calculations – interactions pertinent to formaldehyde hydration. *J. Mol. Struct. (THEOCHEM)*, **88**, pp. 23-35

88. Boys, S.F., Bernardi, F. (1970). The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.*, **19**, pp. 553-566

89. Gillott, S. (1995). Theoretical modelling of molecules related to sialic acid. *Final year project report*, University of Bath.

90. SPARTAN 3.1, Revision A, Wavefunction, Inc., Irvine, CA, 1994

91. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, pp. 926-935

92. Bakowies, D., Thiel, W. (1996). Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.*, **100**, pp. 10580-10594

93. Barnes, J.A., Williams, I.H. (1996). Theoretical modelling of kinetic isotope effects for glycoside hydrolysis in aqueous solution by a hybrid quantum mechanical/molecular mechanical method. *Chem. Commun.*, pp. 193-194

94. Gao, J., Xia, X. (1993). A two-dimensional energy surface for a type II S_N2 reaction in aqueous solution. *J. Am. Chem. Soc.*, **115**, pp. 9667-9675

95. Gao, J. (1996). Hybrid quantum and molecular mechanical simulations: an alternative avenue to solvent effects in organic chemistry. *Acc. Chem. Res.*, **29**, pp. 298-305

96. Gao, J., Pavelites, J.J. (1992). Aqueous basicity of the carboxylate lone pairs and C—O barrier in acetic acid: a combined quantum and statistical mechanical study. *J. Am. Chem. Soc.*, **114**, pp. 1912-1913

97. Gao, J., Xia, X. (1992). A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science*, **258**, pp. 631-635

98. Gao, J. (1992). Absolute free energy of solvation from Monte Carlo simulations using combined quantum and molecular mechanical potentials. *J. Phys. Chem.*, **96**, pp. 537-540

99. Bash, P.A., Ho, L.L., MacKerell, Jr., A.D., Levine, D., Hallstrom, P. (1996). Progress toward chemical accuracy in the computer simulation of condensed phase reactions. *Proc. Natl. Acad. Sci. USA*, **93**, pp. 3698-3703

100. Ho, L.L., MacKerell, Jr., A.D., Bash, P.A. (1996). Proton and hydride transfers in solution: hybrid QM/MM free energy perturbation study. *J. Phys. Chem.*, **100**, pp. 4466-4475

101. Kollman, P. (1993). Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.*, **93**, pp. 2395-2417

102. Connolly, M.L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, **221**, pp. 709-713

103. Moliner, V., Turner, A.J., Williams, I.H. (1997). Transition-state structural refinement with GRACE and CHARMM: realistic modelling of lactate dehydrogenase using a combined quantum/classical method. *Chem. Commun.*, pp. 1271-1272

104. Fields, B.N., Knipe, D.M. (1990). Fields' virology (2^{nd} Ed.). Raven Press, New York

105. Raven, P.H., Johnson, G.B. (1996). Biology. Brown, Iowa

106. Voet, D., Voet, J.G. (1995). Biochemistry (2nd Ed.). Wiley, New York

107. Klenk, H.-D., Rott, R. (1988). The molecular biology of influenza virus pathogenicity. *Adv. Vir. Res.*, **34**, pp. 247-281

108. Darnell, J., Lodish, H., Baltimore, D. (1990). Molecular cell biology (2nd Ed.). Scientific American Books, New York.

109. Klenk, H.-D., Compans, R.W., Choppin, P.W. (1970). An electron microscopic study of the presence or absence of neuraminic acid in enveloped viruses. *Virology*, **42**, pp. 1158-1162

110. Palese, P., Tobita, K., Ueda, M., Compans, R.W. (1974). Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. *Virology*, **61**, pp. 397-410

111. Sigal, G.B., Mammen, M., Dahmann, G., Whitesides, G.M. (1996). Polyacrylamides bearing pendant α-sialoside groups strongly inhibit agglutination of erythrocytes by influenza virus: the strong inhibition reflects enhanced binding through cooperative polyvalent interactions. *J. Am. Chem. Soc.*, **118**, pp. 3789-3799

112. Choi, S.-K., Mammen, M., Whitesides, G.M. (1997). Generation and *in situ* evaluation of libraries of poly(acrylic acid) presenting sialosides as side chains as polyvalent inhibitors of influenza-mediated hemagglutination. *J. Am. Chem. Soc.*, **119**, pp. 4103-4111

113. Wrigley, N.G., Skehel, J.J., Charlwood, P.A., Brand, C.M. (1973). The size and shape of influenza virus neuraminidase. *Virology*, **51**, pp. 525-529

114. Martin, Jr, D.W., Mayes, P.A., Rodwell, V.W., Granner, D.K. (1985). Harper's review of biochemistry (20th ed.). Lange medical publications, Los Altos

115. Varghese, J.N., Laver, W.G., Colman, P.M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9Å resolution. *Nature*, **303**, pp 35-40

116. Varghese, J.N., Colman, P.M. (1991). Three-dimensional structure of the neuraminidase of influenza virus A/Tokyo/3/67 at 2.2Å resolution. *J. Mol. Biol.*, **221**, pp. 473-486

117. Tulip, W.R., Varghese, J.N., Baker, A.T., van Donkelaar, A., Laver, W.G., Webster, R.G., Colman, P.M. (1991). Refined atomic structures of N9 subtype influenza virus neuraminidase and escape mutants. *J. Mol. Biol.*, **221**, pp. 487-497

118. Bossart-Whitaker, P., Carson, M., Bubu, Y.S., Smith, C.D., Laver, W.G., Air, G.M. (1993). Three-dimensional structure of influenza A N9 neuraminidase and its complex with the inhibitor 2-deoxy 2,3-dehydro-N-acetyl neuraminic acid. *J. Mol. Biol.*, **232**, pp. 1069-1083

119. Taylor, G., Garman, E., Webster, R., Saito, T., Laver, G. (1993). Crystallization and preliminary X-ray studies of influenza A virus neuraminidase of subtypes N5, N6, N8 and N9. *J. Mol. Biol.*, **230**, pp. 345-348

120. Burmeister, W.P., Ruigrok, R.W.H., Cusack, S. (1992). The 2.2Å resolution crystal structure of influenza B neuraminidase and its complex with sialic acid. *EMBO Journal*, **11**, pp. 49-56

121. Burmeister, W.P., Henrissat, B., Bosso, C., Cusack, S., Ruigrok, R.W.H. (1993). Influenza B virus neuraminidase can synthesize its own inhibitor. *Structure*, **1**, pp. 19-26

122. von Itzstein, M., Dyason, J.C., Oliver, S.W., White, H.F., Wu, W.-Y., Kok, G.B., Pegg, M.S. (1996). A study of the active site of influenza virus sialidase: an approach to the rational design of novel anti-influenza drugs. *J. Med. Chem.*, **39**, pp. 388-391

123. von Itzstein, M., Wu, W.-Y., Kok, G.B., Pegg, M.S., Dyason, J.C., Jin, B., Phan, T.V., Smythe, M.L., White, H.F., Oliver, S.W., Colman, P.M., Varghese, J.N., Ryan, D.M., Woods, J.M., Bethell, R.C., Hotham, V.J., Cameron, J.M., Penn, C.R. (1993). Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*, **363**, pp. 418-423

124. Chandler, M., Bamford, M.J., Conroy, R., Lamont, B., Patel, B., Patel, V.K., Steeples, I.P., Storer, R., Weir, N.G., Wright, M., Williamson, C. (1995). Synthesis of the potent influenza neuraminidase inhibitor 4-guanidino Neu5ac2en. X-ray molecular structure of 5-acetamido-4-amino-2,6-anhydro-3,4,5-trideoxy-D-*erythro*-L-gluco-nononic acid. *J. Chem. Soc. Perkin Trans. 1*, pp. 1173-1180

125. Smith, P.W., Starkey, I.D., Howes, P.D., Sollis, S.L., Keeling, S.P., Cherry, P.C., von Itzstein, M., Wu, W.Y., Jin, B. (1996). Synthesis and influenza virus sialidase inhibitory activity of analogues of 4-guanidino-Neu5ac2en (GG167) with modified 5-substituents. *Eur. J. Med. Chem.*, **31**, pp. 143-150

126. Starkey, I.D., Mahmoudian, M., Noble, D. Smith, P.W., Cherry, P.C., Howes, P.D., Sollis, S.L. (1995). Synthesis and influenza virus sialidase inhibitory activity of the 5-desacetamido analogue of 2,3-didehydro-2,4-dideoxy-4-guanidinyl-N-acetylneuraminic acid (GG167). *Tetrahedron Lett.*, **36**, pp. 299-302

127. Bamford, M.J., Pichel, J.C., Husman, W., Patel, B., Storer, R., Weir, N.G. (1995). Synthesis of 6-, 7- and 8-carbon sugar analogues of potent anti-influenza 2,3-didehydro-2,3-dideoxy-N-acetylneuraminic acid derivatives. *J. Chem. Soc. Perkin Trans. 1*, pp. 1181-1187

128. Chandler, M., Conroy, R., Cooper, A.W.J., Lamont, R.B., Scicinski, J.J., Smart, J.E., Storer, R., Weir, N.G., Wilson, R.D., Wyatt, P.G. (1995). Approaches to carbocyclic analogues of the potent neuraminidase inhibitor 4-guanidino-Neu5ac2en. X-ray molecular structure of N-[(1S,2S,6R)-2-azido-6-benzyloxymethyl-4-formylcyclohex-3-enyl] acetamide. *J. Chem. Soc. Perkin Trans. 1*, pp. 1189-1197

129. Hayden, F.G., Osterhaus, A.D.M.E., Treanor, J.J., Fleming, D.M., Aoki, F.Y., Nicholson, K.G., Bohnen, A.M., Hirst, H.M., Keene, O., Wightman, K. (1997). Efficacy and safety of the neuraminidase inhibitor zanamivir in the treatment of influenzavirus infections. *New. Eng. J. Med.*, **337**, pp. 874-880

130. For current information on Relenza™ (zanamivir), visit the GlaxoWellcome web site at *http://www.glaxowellcome.co.uk* (correct address at time of writing).

131. Kim, C.U., Lew, W., Williams, M.A., Liu, H., Zhang, L., Swaminathan, S., Bischofberger, N., Chen, M.S., Mendel, D.B., Tai, C.Y., Laver, W.G., Stevens, R.G. (1997). Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.*, **119**, pp. 681-690

132. Lew, W., Williams, M.A., Mendel, D.B., Escarpe, P.A., Kim, C.U. (1997). C_3-thia and C_3-carba isosteres of a carbocyclic influenza neuraminidase inhibitor, (3R,4R,5S)-4-acetamido-5-amino-3-propoxy-1-cyclohexene-1-carboxylic acid. *Bioorg. Med. Chem. Lett.*, **7**, pp. 1843-1846

133. Williams, M.A., Lew, W., Mendel, D.B., Tai, C.Y., Escarpe, P.A., Laver, W.G., Stevens, R.C., Kim, C.U. (1997). Structure-activity relationships of carbocyclic influenza neuraminidase inhibitors. *Bioorg. Med. Chem. Lett.*, **7**, pp. 1837-1842

134. Zhang, L., Williams, M.A., Mendel, D.B., Escarpe, P.A., Kim, C.U. (1997). Synthesis and activity of C_2-substituted analogs of influenza neuraminidase inhibitor GS 4071. *Bioorg. Med. Chem. Lett.*, **7**, pp. 1847-1850

135. Smith, P.W., Robinson, J.E., Evans,D.N., Sollis, S.L., Howes, P.D., Trivedi, N., Bethell, R.C. (1999). Sialidase inhibitors related to zanamivir: synthesis and biological evaluation of 4H-pyran 6-ether and ketone. *Bioorg. Med. Chem. Lett.*, **9**, pp. 601-604

136. Smith P.W., Sollis, S.L., Howes, P.D., Cherry, P.C., Starkey, I.D., Cobley, K.N., Weston, H., Scicinski, J., Merritt, A., Whittington, A., Wyatt, P., Taylor, N., Green, D., Bethell, R., Madar, S., Fenton, R.J., Morley, P.J., Pateman, T., Beresford, A. (1998). Dihydropyrancarboxamides related to zanamivir: a new series of inhibitors of influenza sialidases. 1. Discovery, synthesis, biological activity and structure-activity relationships of 4-guanidino- and 4-amino-4H-pyran-6-carboxamides. *J. Med. Chem.*, **41**, pp. 787-797

137. Smith P.W., Sollis, S.L., Howes, P.D., Cherry, P.C., Cobley, K.N., Taylor, H., Whittington, A.R., Scicinski, J., Bethell, R.C., Taylor, N., Skarzynski, T., Cleasby, A., Singh, O., Wonacott, A., Varghese, J., Colman, P. (1996). Novel inhibitors of influenza sialidases related to GG167. Structure-activity, crystallographic and molecular dynamics studies with 4H-pyran-2-carboxylic acid 6-carboxamides. *Bioorg. Med. Chem. Lett.* **6**, pp. 2931-2936

138. Sollis, S.L., Smith, P.W., Howes, P.D., Cherry, P.C., Bethell, R.C. (1996). Novel inhibitors of influenza sialidase related to GG167. Synthesis of 4-amino and guanidino-4H-pyran-2-carboxylic acid-6-propylamides; selective inhibitors of influenza A virus sialidase. *Bioorg. Med. Chem. Lett.*, **6**, pp. 1805-1808

139. Taylor, N.R., Cleasby A., Singh, O., Skarzynski, T., Wonacott, A.J., Smith, P.W., Sollis, S.L., Howes, P.D., Cherry, P.C., Bethell, R., Colman, P., Varghese, J. (1998). Dihydropyrancarboxamides related to zanamivir: a new

series of inhibitors of influenza virus sialidases. 2. Crystallographic and molecular modeling study of complexes of 4-amino-4H-pyran-6-carboxamides and sialidase from influenza virus types A and B. *J. Med. Chem.* **41**, pp.798-807

140. Williams, M., Bischofberger, N., Swaminathan, S., Kim, C.U. (1995). Synthesis and influenza neuraminidase inhibitory activity of aromatic analogues of sialic acid. *Bioorg. Med. Chem. Lett.*, **5**, pp. 2251-2254

141. Singh, S., Jedrzejas, M.J., Air, G.M., Luo, M., Laver, W.G., Brouillette, W.J. (1995). Structure-based inhibitors of influenza virus sialidase. A benzoic acid lead with novel interaction. *J. Med. Chem.*, **38**, pp. 3217-3225

142. Jedrzejas, M.J., Singh, S., Brouillette, W.J., Laver, W.G., Air, G.M., Luo, M. (1995). Structures of aromatic inhibitors of influenza virus neuraminidase. *Biochemistry*, **34**, pp. 3144-3151

143. Chand, P., Babu, Y.S., Bantia, S., Chu, N., Cole, L.B., Kotian, P.L., Laver, W.G., Montgomery, J.A., Pathak, V.P., Petty, S.L., Shrout, D.P., Walsh, D.A., Walsh, G.M. (1997). Design and synthesis of benzoic acid derivatives as influenza neuraminidase inhibitors using structure-based drug design. *J. Med. Chem.*, **40**, pp. 4030-4052.

144. Sudbeck, E.A., Jedrzejas, M.J., Singh, S., Brouillette, W.J., Air, G.M., Laver, W.G., Babu, Y.S., Bantia, S., Chand, P., Chu, N., Montgomery, J.A., Walsh, D.A., Luo, M. (1997). Guanidinobenzoic acid inhibitors of influenza virus neuraminidase. *J. Mol. Biol.*, **267**, pp. 584-594.

145. Blick, T.J., Tiong T., Sahasrabudhe, A., Varghese, J.N., Coleman, P.M., Hart, G.J., Bethell, R.C., McKimm-Breschkin, J.L. (1995). Generation and characterization of an influenza virus neuraminidase variant with decreased sensitivity to the neuraminidase-specific inhibitor 4-guanidino-Neu5ac2en. *Virology*, **214**, pp. 475-484

146. Staschke, K.A., Colacino, J.M., Baxter, A.J., Air, G.M., Bansal, A., Hornback, W.J., Munroe, J.E., Laver, W.G. (1995). Molecular basis for the resistance of influenza viruses to 4-guanidino-Neu5ac2en. *Virology*, **214**, pp. 642-646

147. Pauling, L. (1948). Nature of forces between large molecules of biological interest. *Nature*, **161**, pp. 707-709

148. Wolfenden, R. (1969). Transition state analogues for enzyme catalysis. *Nature*, **223**, pp. 704-705

149. Warshel, A., Levitt, M. (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, **103**, pp. 227-249

150. Phillips, D.C. (1967). The three-dimensional structure of an enzyme molecule. *Sci. Am.*, **215**, 5, pp. 78-90

151. Waszkowycz, B., Hillier, I.H., Gensmantel, N., Payling D.W. (1991). A combined quantum-mechanical molecular mechanical model of the potential-energy surface of ester hydrolysis enzyme phospholipase-A_2. *J. Chem. Soc. Perkin Trans. 2*, pp. 225-231

152. Waszkowycz, B., Hillier, I.H., Gensmantel, N., Payling D.W. (1991). Combined quantum mechanical-molecular mechanical study of catalysis by the enzyme phospholipase A_2: an investigation of the potential energy surface for amide hydrolysis. *J. Chem. Soc. Perkin Trans. 2.*, pp. 2025-2032

153. Bash, P.A., Field, M.J., Davenport, R.C., Petsko, G.A., Ringe, D., Karplus, M. (1991). Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry*, **30**, pp. 5826-5832

154. Lyne, P.D., Mulholland, A.J., Richards W.G. (1995). Insights into chorismate mutase catalysis from a combined QM/MM simulation of the enzyme reaction. *J. Am. Chem. Soc.*, **117**, pp. 11345-11350

155. Mulholland, A.J., Richards, W.G. (1997). Acetyl-CoA enolization in citrate synthase: a quantum mechanical/molecular mechanical (QM/MM) study. *Proteins*, **27**, pp. 9-25

156. Harrison M.J., Burton N.A., Hillier, I.H., Gould, I.R. (1996). Mechanism and transition state structure for papain catalysed amide hydrolysis, using a hybrid QM/MM potential. *Chem. Commun.*, pp. 2769-2770

157. Harrison M.J., Burton N.A., Hillier, I.H. (1997). Catalytic mechanism of the enzyme papain: predictions with a hybrid quantum mechanical/molecular mechanical potential. *J. Am. Chem. Soc.*, **119**, pp. 12285-12291

158. Pearlman, D.A., Case, D.A., Caldwell, J.C., Seibel, G.L., Singh, U.C., Weiner, P., Kollman, P.A., AMBER 4.0, University of California, San Francisco, 1994.

159. Turner, A.J., Moliner, V., Williams, I.H. (1999). Transition-state structural refinement with GRACE and CHARMM: flexible QM/MM modelling for lactate dehydrogenase. *Phys. Chem. Chem. Phys.*, **1**, pp. 1323-1331

160. Ranganathan, S., Gready, J.E. (1997). Hybrid quantum and molecular mechanical (QM/MM) studies on pyruvate to L-lactate interconversion in L-lactate dehydrogenase. *J. Phys. Chem. B.*, **101**, pp. 5614-5618

161. Cunningham, M.A., Ho, L.L., Nguyen, D.T., Gillilan, R.E., Bash, P.A. (1997). Simulation of the enzyme reaction mechanism of malate dehydrogenase. *Biochemistry*, **36**, pp. 4800-4816

162. Cummins, P.L., Gready, J.E. (1998). Molecular dynamics and free energy perturbation study of hydride-ion transfer step in dihydrofolate reductase

using combined quantum and molecular mechanical model. *J. Comput. Chem.*, **19**, pp. 977-988

163. Jourand, D. (1997). Étude Théorique du Mécanisme Réactionnel de la Neuraminidase d'influenza B. *Ph.D. thesis*, L'universite Joseph Fourier-Grenoble-I.

164. For the purposes of discussion in this thesis, the residue side chain numbers refer to the designation given in influenza A sialidase (N2). This is to avoid confusion. However, for information Asp-151 = Asp-149, Glu-277 = Glu-275 and Tyr-406 = Tyr-408 in the influenza B sialidase.

165. Barnes, J.A., Williams, I.H. (1996). Quantum mechanical/molecular mechanical approaches to transition state structure: mechanism of sialidase action. *Biochem. Soc. Trans.*, **24**, pp. 263-268

166. Barnes, J.A. (1994). Theoretical modelling of transition states for chemical processes. *Ph.D. thesis*, University of Bath.

167. Chong, A.K.J., Pegg, M.S., Taylor, N.R., von Itzstein, M. (1992). Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus. *Eur. J. Biochem.*, **207**, pp. 335-343

168. Tiralongo, J., Pegg, M.S., von Itzstein, M. (1995). Effect of substrate aglycon on enzyme mechanism on the reaction of sialidase from influenza virus. *FEBS Lett.*, **372**, pp. 148-150

169. Janakiraman M.N., White C.L., Laver, W.G., Air, G.M., Luo, M. (1994). Structure of influenza virus neuraminidase B/Lee/40 complexed with sialic acid and a dehydro analog at 1.8-Å resolution: implications for the catalytic mechanism. *Biochemistry*, **33**, pp. 8172-8179

170. Smith, B.J. (1997). A conformational study of 2-oxonol: insight into the role of ring distortion on enzyme-catalyzed glycoside bond cleavage. *J. Am. Chem. Soc.*, **119**, pp. 2699-2706

171. Dewar, M.J.S., Dieter K.M. (1986). Evaluation of AM1 calculated proton affinities and deprotonation enthalpies. *J. Am. Chem. Soc.*, **108**, pp. 8075-8086

172. Voets, R., François, J.–P., Martin, J.M.L., Mullens, J., Yperman, J., van Poucke, L.C. (1990). Theoretical study of the proton affinities of 2-, 3- and 4-monosubstituted phenolate ions in the gas phase by means of MINDO/3, MNDO, and AM1. *J. Comput. Chem.*, **11**, pp. 269-290

173. Morpurgo, S., Bossa, M., Morpurgo, G.O. (1998). Critical test of PM3-calculated proton transfer activation energies: a comparison with *ab initio* and AM1 calculations. *J. Mol. Struct. (THEOCHEM)*, **429**, pp. 71-80

174. Lide, D.R. (1996), Handbook of physics and chemistry (77th Ed). *CRC Press*

175. Lapthorn, C.L. (1998). Low barrier hydrogen bonds and enzymatic catalysis and theoretical modelling of two residue pairs in the sialidase enzyme. *Final year project report*, University of Bath

176. Alex, A., Finn, P. (1997). Fast and accurate predictions of relative binding energies. *J. Mol. Struct.(THEOCHEM)*, **398-399**, pp. 551-554

APPENDICES

Appendix 1: Example CHARMM Scripts

(a). Script for Explicit Solvation QM/MM Studies

This script represents a basic optimisation of a solute immersed within a 15Å water sphere. The specifics relate to the work in chapter 6, where the solvent sphere is divided into *immediate* and *bulk* solvent. The simulated annealing part of the script was only done on the *immediate* solvent molecules surrounding the fixed solute. The IC constraint commands used to calculate the reaction profile for the full system are shown in appendix 1b.

The first stage is to initialise all the variables. This is not essential, but it is good programming practice. `Bomb -1` instructs CHARMM to run the script unless a severe problem (e.g. missing MM parameters) is encountered.

```
set 1 empty
set 2 empty
set 3 empty
set 4 empty
set 5 empty
set 6 empty
set 7 empty
set 8 empty
set 9 empty
set a empty
set b empty
set c empty
set d empty
set e empty
set f empty
set g empty
set h empty
set i empty
set j empty
set k empty
set l empty
set m empty
set n empty
set p empty
set q empty
bomb -1
```

Open the topology, parameter and co-ordinate files; read in the topology and parameter information. The topology file contains information on the connectivity of defined residues, whereas the parameter file contains the data required to calculate the MM energy of the system. Each file is associated with a "unit" and is read as a

formatted PDB file (FORM) or as a CARD file; "!" represents a comment. The sequence input gives information on the residues used to build the molecular system.

```
OPEN NAME ./prot_na_bill_top_mk2flu.inp          UNIT 11 CARD READ
OPEN NAME ./par_prot_na_bill.inp                 UNIT 12 CARD READ
OPEN NAME ./las10md40kc05mk4.pdb                 UNIT 13 FORM READ
!OPEN NAME ./outer5.pdb                          UNIT 14 FORM READ
OPEN NAME ./lac_seq.inp                          UNIT 15 CARD READ
READ  RTF CARD                                   UNIT 11
READ  PARAMETER CARD                             UNIT 12
```

The molecular system must now be defined in terms of residues; from this information, arrays are generated to store information about the system. In this case, the solute (α-lactone) is identified by the segment NEU5. This is the identifier used for all sialyl glycoside moieties modelled in this thesis. FSOL represents the 140 water molecules in the *immediate* environment of the solute and SOLV identifies 345 solvent molecules in the *bulk* environment. This present script then reads in the co-ordinates of the structure (chapter 6).

```
READ SEQUENCE CARD                                   UNIT 15
GENERATE neu5 SETUP

READ SEQUENCE TIP3 140
GENERATE fsol SETUP NOANG NODIHE

READ SEQUENCE TIP3 345
GENERATE solv SETUP NOANG NODIHE


READ COOR PDB                                        UNIT 13
```

Now set the non-bonded atom-based cut-offs to very large (40Å) and apply the solvent boundary potential to maintain the 15Å water sphere. This is applied (in this case) to oxygen atoms in the *bulk* solvent. During the generation of the different *immediate* solvent structures using simulated annealing, the 10Å boundary potential was applied to the oxygen atoms of the FSOL water molecules. "-" allows the continuation of a command onto the next line.

```
nbonds atom switch cutnb 40.00 ctofnb 38 ctonnb 36

OPEN NAME  ./wat15.pot UNIT 4 FORM READ
sbound READ UNIT 4
sboundary set xref 0.0 yref 0.0 zref 0.0 -
```

```
assign 1 selection (segid solv .and. type O*) end
```

After setting up the system, it is now possible to restart a simulation if required. This is done by reading the restart file into the comparison co-ordinate array and then swapping these with the co-ordinates of the current system. This ensures that all the atoms in the system are precisely set up as in the original simulation. However, this stage is not always necessary and is thus commented out until needed.

```
!open name t4s11solvic_min.pdb unit 77 form read
!read coor comp pdb unit 77
!close unit 77
!coor swap
```

Select the solute to be treated with the AM1 Hamiltonian; the overall charge of the system is zero. This line is commented out when using an MM only potential.

```
QUANTUM SELECT SEGID neu5 END AM1 CHARGE 0
```

The script up to this point will be defined as the *CHARMM QM/MM system set up.*

Sometimes some rigid constraints may need to be applied. In adding the *bulk* solvent sphere to the solute and *immediate* solvent sphere, both `NEU5` and `FSOL` segments were kept fixed.

```
CONS FIX SELECT (SEGID neu5 .OR. SEGID fsol) END
```

Next, a file called `arguments` is generated containing lines within a title (such lines start with a "*"). The title reports the type of optimiser used, the gradient tolerance (in kcal mol^{-1} $Å^{-1}$), the maximum allowed optimisation cycles, the restart file name and the number of cycles allowed before each restart file is generated. All titles in CHARMM must finish with a line just containing an "*".

```
OPEN NAME arguments                        UNIT  77 FORM WRITE
WRITE TITLE UNIT 77
* set 1 abnr                  ! select ABNR minimiser
* set 2 0.005                 ! select GRMS exit criterion
* set 5 600                   ! select the maximum number of cycles
* set 6 las10md40kc05mk5.res  ! select the name of the restart file
```

```
* set 7 0                          ! Number of cycles done
* set 8 200                        ! Set the required number of cycles
* return
*
```

Next, the variables, `r` and `s` are used to control the loops that occur later in the script. The `goto` command first goes to the MD part of the script before the final minimisation of the system is done.

```
set r main
set s dyn_run
goto DYNAMICS
label dyn_run
goto MINIMISE
label main
```

The command, `STOP`, is used to finish the simulation and to terminate the running of CHARMM.

```
STOP
```

The first listed procedure is minimisation. The label sets the position in the script recognised by the `goto` command; "`goto MINIMISE`" will thus initiate a minimisation of the system. The first task for this process involves reading in the arguments; the `return` command in the `Arguments` title instructs CHARMM to return to the input script.

```
label MINIMISE

OPEN NAME arguments                                UNIT 77 FORM READ
STREAM UNIT 77
```

The next stage is to minimise the system using (in this case) the ABNR optimiser for 200 steps with an initial step size of 0.02Å. System energies are printed after 20 cycles. The 200 steps are fully carried out before the PDB restart file is written; the total number of steps done (variable 7) is then increased by this number. The restart file title gives information on the total energy and gradient of the system and the number of ABNR cycles already done. Unit 77 is then closed off for future use.

```
mini @1 nstep @8 nprint 20 step 0.02 tolgrd 0.0
incr 7 by @8

OPEN NAME @6                                    UNIT 77 FORM WRITE
WRITE COOR PDB UNIT 77
* SOLVATED Neu5ac LA in 15A sphere
* THIS OPTIMISATION:
* @7 cycles @1
* Total Energy is ?ENER with Gradient ?GRMS
*
CLOSE UNIT 77
```

Now the RMS gradient tolerance is tested. If it is below the specified value (0.005 kcal mol^{-1} $Å^{-1}$) then the script moves forward to `minimiser_gohome` which sends the program back to the `main` label (just before the `STOP` command) in the script. However, if the gradient is too high, the program is sent back to the `minimise_loop` label for a further 200 optimisations cycles. Applying the gradient tolerance test in this way (rather than specifying it when minimising the system) reduces the chances of premature terminations in the minimisation procedure. "Premature terminations" can sometimes occur because the RMS gradient of the (large) system tends to "bounce"; such an effect can sometimes drop the RMS gradient below the tolerance before the system is fully minimised. The script also moves to `minimiser_gohome` if the maximum number of optimisation steps has been taken.

```
if 7 eq @5 goto minimiser_gohome
if 2 gt ?GRMS goto minimiser_gohome
goto minimise_loop

label minimiser_gohome
goto @r
```

The next label initiates the molecular dynamics part of the script.

```
label DYNAMICS
```

The first stage is to briefly optimise the system to avoid uncontrolled KE increase during the dynamics run.

```
MINI ABNR NSTEP 10000 STEP 0.02 tolgrd 0.05 nprint 100
```

The second stage is to heat and equilibrate the system using the Leapfrog algorithm. This is done over 40ps (NSTEP * TIMESTEP) using a time step (TIMESTEP) of 1fs. Data is written to the output file every 100fs (NPRINT), with average properties and fluctuations printed every 1000fs (IPRFRQ). The temperature is raised from 0K (FIRSTT) to 300K (FINALTT); this is done in 10K increments (TEMINC) with 200fs separating each temperature increment (IHTFREQ). A temperature window of ±10K is maintained (TWINDH, TWINDL).

```
DYNAMICS LEAP VERLET STRT  NSTEP 40000 TIMESTEP 0.001 -
    IPRFRQ 1000 IHTFRQ 200 IEQFRQ 1000 NTRFRQ 1000  -
    IUNREA -1 IUNWRI -1 IUNCRD -1 IUNVEL -1 KUNIT 70 -
    NPRINT 100 NSAVC 100 NSAVV 0 INBFRQ -1  -
    FIRSTT 0.0 FINALT 300.0 TEMINC 10.0   -
    IASORS 1 IASVEL 1 ISCVEL 0 ICHECW 1 TWINDH 10.0 TWINDL -10.0
```

The final stage is to cool the system from 300K down to absolute zero. This is done over 160ps with a time step of 1fs. The temperature is initially decreased by 0.5K every 200fs for 120ps. Cooling continues for a further 40ps since the temperature of the system will still not be at absolute zero. No temperature window is applied here. Information about the system is printed every 100fs, with fluctuations and averages printed every 1000fs.

```
DYNAMICS LEAP VERLET STRT  NSTEP 160000 TIMESTEP 0.001 -
    IPRFRQ 1000 IHTFRQ 200 IEQFRQ 1000 NTRFRQ 1000  -
    IUNREA -1 IUNWRI -1 IUNCRD -1 IUNVEL -1 KUNIT 70 -
    NPRINT 100 NSAVC 50 NSAVV 0 IHBFRQ 0 INBFRQ -1  -
    TSTRUC 300.0 FIRSTT 300.0 FINALT 0.0 TEMINC -0.5   -
    IASORS 0 IASVEL 0 ISCVEL 0 ICHECW 1 TWINDH 0.0
```

Following simulated annealing, CHARMM returns to the main script (at the dyn_run label).

```
goto @s
```

(b). Script for Enzyme QM/MM Studies

The first stage initialises all the variables, opens all the relevant files and reads in the topology and parameter data. The syntax for doing this is the same as in appendix 1a and so is not reported here.

The structure file is now built. Firstly, the substrate is generated and designated as segment NEU5. The enzyme is then built using the amino acid sequence in appendix 1c (shown in the header of 2BAT). The crystallographic determined water molecules are then set up and identified as WATR; the calcium ion bound to the protein is then generated with identifier CION. Finally, the co-ordinate file for the ES complex is then read.

```
READ SEQUENCE CARD                                          UNIT 16
GENERATE neu5 SETUP

READ SEQUENCE CARD                                          UNIT 15
GENERATE prot SETUP

READ SEQUENCE TIP3 116
GENERATE watr SETUP NOANG NODIHE

READ SEQUENCE CARD                                          UNIT 17
GENERATE cion SETUP

READ COOR PDB                                               UNIT 13
```

Because there are disulfide links in the NA structure, CHARMM must be instructed to delete the —SH hydrogen atoms and form the S—S links between the cysteine residues.

```
PATCH DISU prot 94  prot 112 SETUP
PATCH DISU prot 197 prot 210 SETUP
PATCH DISU prot 199 prot 208 SETUP
```

A 17Å water sphere is then added to the system; the 694 added solvent molecules are designated SOLV.

```
READ SEQUENCE TIP3 694
GENERATE solv SETUP NOANG NODIHE
      READ COOR PDB APPEND                                  UNIT 14
```

The QM link atom is then added to the Asp-151 side chain. To incorporate it into the "start-up" system, it is placed half way between the C_α—C_β bond axis. The QM link atom is known as HQA.

```
COOR AXIS SELECT ( (SEGID prot .AND. RESID 70 ) .AND. TYPE CA) END -
SELECT ((SEGID prot .AND. RESID 70 ) .AND. TYPE CB) END

COOR SET xdir ?xcen ydir ?ycen zdir ?zcen SELECT (SEGID prot .AND. -
RESID 70 .AND. TYPE HQA) END
```

The next stage is to remove unwanted atoms from the system. Protein and WATR residues beyond 15Å of the substrate are deleted. A SOLV molecule whose oxygen is within 2.5Å of any heavy atom in the ES complex is also deleted.

```
DELETE ATOM SELECT .NOT. (.BYRES. (SEGID neu5 .AROUND. 15.00) .OR. -
(SEGID solv)) END

DELETE ATOM SELECT (.BYRES. ( (SEGID solv .AND. TYPE OH2 ) .AND.-
 ( (.NOT. SEGID solv .AND. .NOT. HYDROGEN) .AROUND. 2.5) ) ) END
```

Atom-based non-bonded cut-offs are set to very large (40Å) and a 19Å boundary potential is applied to the SOLV and WATR molecules. For heavy protein and crystallographic water atoms beyond 13Å of the origin, a harmonic force constraint is applied. The force constant is mass-weighted with a value of 1.0.

```
nbonds atom switch cutnb 40.00 ctofnb 38 ctonnb 36
```

```
OPEN NAME  ./wat19.pot UNIT 4 FORM READ
sbound READ UNIT 4
sboundary set xref 0.0 yref 0.0 zref 0.0 -
          assign 1 selection (segid watr .and. type O*) .OR.-
(segid solv .and. type O*) end

CONSTRAINT HARMONIC FORCE 1.0 SELECT ((SEGID prot .OR. SEGID watr) -
 .AND.  (.NOT. TYPE H*) .AND. (.NOT. POINT 0.0 0.0 0.0 CUT 13.0))-
 END  MASS EXPONENT=2
```

After setting up the system, it is now possible to restart a simulation if required. This is done by reading the restart file into the comparison co-ordinate array and swapping it with the co-ordinates of the current system. This ensures that all the atoms in the restart file are set up as in the original simulation. The QM link atom in

the restart file will have already been added; initially such an atom is placed 1.16Å from the C_β atom.

```
open name fluptr1s7mk2pt3.pdb unit 77 form read
read coor comp pdb unit 77
close unit 77
coor swap
```

It is often useful to define a sub-set of atoms. The QM link atom is included in the `QuantumASP` definition.

```
DEFINE QuantumASP SELECT -
(RESID 70 .AND. (TYPE CB .OR. TYPE HB1 .OR. TYPE HB2 .OR. TYPE CG -
.OR. TYPE OD2 .OR. TYPE OD1 .OR. TYPE HD2 .OR. TYPE HQA) ) END

DEFINE MolMechSolv SELECT ( (SEGID solv) .OR. (SEGID watr .AND. -
.NOT. (RESID 146 .OR. RESID 148) ) ) END

DEFINE pnp SELECT ( (SEGID neu5) .AND. (TYPE C19 .OR. TYPE C20.OR. -
TYPE C21 .OR. TYPE C22 .OR. TYPE N23 .OR. TYPE C24 .OR. TYPE C25 -
.OR. TYPE H45 .OR. TYPE H46 .OR. TYPE O47 .OR. TYPE O48 .OR. TYPE-
 H49 .OR. TYPE H50) ) END
```

The initial adiabatic mapping process can be achieved in at least two ways. The first way is to define a vector for a transferring atom. This was done in the enzyme study when investigating the proton transfer and attack of water to the zwitterion. The translation vector for the proton transfer was calculated from the proton's position in the reactant and its position in the intermediate. The vector is then divided into roughly ten sections (in Cartesian co-ordinates) and the atom is translated by this amount to produce the energy profile. The `Quick` command can give the distance between two atoms (in this case 38 and 40) and thus serves as a check to see if the translation has correctly occurred. The second approach was used in the α-lactone/zwitterion solvation study. This involves extending the bond to a specific value and then applying an internal co-ordinate constraint. This approach is included here but is commented out.

```
SET a -0.06
SET b 0.065
SET c -0.0783
COOR TRANS SELECT (RESID 70 .AND. TYPE HD2) END xdir @a ydir @b -
zdir @c
```

```
QUICK 38 40

!set dist1 ?DIST
!incr dist1 by 0.05
!IC DELETE SELECT .NOT. (BYNUM 38 .OR. BYNUM 40) END
!IC EDIT
!BOND BYNU 38 BYNU 40 @dist1
!END
!CONS IC BOND 1e7 EXPO 2
```

The substrate and Asp(H)-151 side chain are treated with the AM1 Hamiltonian. Because the Asp residue is protonated, the overall charge of the QM system is –1 (associated with the carboxylate group of the substrate).

```
QUANTUM SELECT (SEGID neu5 .OR. QuantumASP) END AM1 CHARGE -1
```

The script up to this point is defined as the *CHARMM QM/MM system set up.*

Rigid constraints are now applied. For the enzyme system, this will include the QM link atom. To generate the reaction profile, the proton on the Asp(H)-151 carboxylate group and the corresponding carboxylate oxygen are also fixed.

```
CONS FIX SELECT (TYPE HQA .OR. (RESID 70 .AND. (TYPE OD2 .OR. TYPE -
 HD2))) END
```

A file called `arguments` is generated containing lines within a title (such lines start with a "*"). These report the type of optimiser used, the gradient tolerance (in kcal mol^{-1} $Å^{-1}$), the maximum allowed optimisation cycles, the restart file name, the number of optimisation cycles already done and the number of cycles allowed before each that restart file is generated.

```
OPEN NAME arguments                          UNIT  77 FORM WRITE
WRITE TITLE UNIT 77
* set 1 abnr                  ! select ABNR minimiser
* set 2 0.005                 ! select GRMS exit criterion
* set 5 600                   ! select the maximum number of cycles
* set 6 fluptr2s8.res         ! select the name of the restart file
* set 7 0                     ! number of cycles already done
* set 8 100                   ! Set the required number of cycles
* return
*
CLOSE UNIT 77
```

The variable, r, is used to control the loops that occur later in the script. The goto command goes to the minimisation part of the script.

```
set r min_home_2
goto MINIMISE
label min_home_2
```

The command, STOP, is used to finish the simulation and to terminate the running of CHARMM.

```
STOP
```

The label sets the position of the script recognised by the goto command; "goto MINIMISE" will thus initiate a minimisation of the system. The first stage of minimisation using this script is to read in the arguments.

```
label MINIMISE

OPEN NAME arguments                    UNIT 77 FORM READ
STREAM UNIT 77
```

Minimise the system using the ABNR optimiser for 600 steps with an initial step size of 0.02Å; system energies are printed after 20 cycles. The 600 steps are fully carried out before the total-number-of-steps-done counter is then increased by 600. The restart file title gives information on the total energy and RMS gradient of the system, and the number of ABNR cycles already done. It also includes the current length of the bond being elongated. The title may also contain a history of previous optimisations done to arrive at the restart structure. Unit 77 is then closed off for future use.

```
label minimise_loop

mini @1 nstep @8 nprint 20 step 0.02 tolgrd 0.0
incr 7 by @8

OPEN NAME @6                                    UNIT 77 FORM WRITE
WRITE COOR PDB UNIT 77
* PREVIOUS OPTIMISATIONS:
*   As For SIA and QM/MM Quantised (flusi0qmmm) and
*   QM/MM 800 cycles ABNR SIP QM ELSE MM; H20 FIXED
```

```
*   QM/MM 3000 cycles ABNR SIP/ASP151 QM ELSE MM; H20 FIXED
* THIS OPTIMISATION:
*     @7 cycles @1
*       ASP-151 H20 DISTANCE IS @9
* created by STUART FIRTH-CLARK, UNIVERSITY OF BATH
*
CLOSE UNIT 77
```

Now the gradient tolerance is tested. If it is below the specified value (0.005 kcal mol^{-1} $Å^{-1}$) then the script moves forward to `minimiser_gohome` which sends the program back to the `main` label (just before the `STOP` command) in the script. However, if the gradient is too high, the program is sent back to the `minimise_loop` label for a further 100 optimisations cycles. Applying the gradient tolerance test in this way (rather than specifying it when minimising the system) reduces the chances of premature terminations in the minimisation procedure. "Premature terminations" can sometimes occur because the RMS gradient of the (large) system "bounces"; such an effect can sometimes drop the RMS gradient below the tolerance before the system is fully minimised. The script also moves forward if the number of optimisation cycles has reached the maximum value (given by variable 5)

```
if 7 eq @5 goto minimiser_gohome
if 2 gt ?GRMS goto minimiser_gohome
goto minimise_loop

label minimiser_gohome

goto @r
```

(c). Amino acid sequence for influenza sialidase

Sequence files begin with a title and the number of residues in the sequence. The list of residues then follows.

```
*Influenza-NA sequence
*
388
VAL GLU TYR ARG ASN TRP SER LYS PRO GLN CYS GLN ILE
THR GLY PHE ALA PRO PHE SER LYS ASP ASN SER ILE ARG
LEU SER ALA GLY GLY ASP ILE TRP VAL THR ARG GLU PRO
TYR VAL SER CYS ASP PRO VAL LYS CYS TYR GLN PHE ALA
LEU GLY GLN GLY THR THR LEU ASP ASN LYS HIS SER ASN
ASP THR VAL HIS ASPQ ARG ILE PRO HIS ARG THR LEU LEU
MET ASN GLU LEU GLY VAL PRO PHE HIS LEU GLY THR ARG
GLN VAL CYS ILE ALA TRP SER SER SER SER CYS HIS ASP
GLY LYS ALA TRP LEU HIS VAL CYS ILE THR GLY ASP ASP
LYS ASN ALA THR ALA SER PHE ILE TYR ASP GLY ARG LEU
VAL ASP SER ILE GLY SER TRP SER GLN ASN ILE LEU ARG
THR GLN GLU SER GLU CYS VAL CYS ILE ASN GLY THR CYS
THR VAL VAL MET THR ASP GLY SER ALA SER GLY ARG ALA
ASP THR ARG ILE LEU PHE ILE GLU GLU GLY LYS ILE VAL
HIS ILE SER PRO LEU ALA GLY SER ALA GLN HIS VAL GLU
GLU CYS SER CYS TYR PRO ARG TYR PRO GLY VAL ARG CYS
ILE CYS ARG ASP ASN TRP LYS GLY SER ASN ARG PRO VAL
VAL ASP ILE ASN MET GLU ASP TYR SER ILE ASP SER SER
TYR VAL CYS SER GLY LEU VAL GLY ASP THR PRO ARG ASN
ASP ASP ARG SER SER ASN SER ASN CYS ARG ASN PRO ASN
ASN GLU ARG GLY THR GLN GLY VAL LYS GLY TRP ALA PHE
ASP ASN GLY ASN ASP LEU TRP MET GLY ARG THR ILE SER
LYS ASP LEU ARG SER GLY TYR GLU THR PHE LYS VAL ILE
GLY GLY TRP SER THR PRO ASN SER LYS SER GLN ILE ASN
ARG GLN VAL ILE VAL ASP SER ASP ASN ARG SER GLY TYR
SER GLY ILE PHE SER VAL GLU GLY LYS SER CYS ILE ASN
ARG CYS PHE TYR VAL GLU LEU ILE ARG GLY ARG LYS GLN
GLU THR ARG VAL TRP TRP THR SER ASN SER ILE VAL VAL
PHE CYS GLY THR SER GLY THR TYR GLY THR GLY SER TRP
PRO ASP GLY ALA ASN ILE ASN PHE MET PRO ILE
```

Appendix 2: Example GRACE Scripts

The basic GRACE program (coded in Tool Control Language (TCL)) doesn't evaluate energies or structures, but controls the operation of CHARMM which does. In addition, GRACE also includes a FORTRAN linear algebra routine (`la`) that is interfaced with CHARMM through the TCL part of the program. `La` contains the extra functionality required when modelling chemical reactivity using the CHARMM QM/MM potential.

(a). Script for Transition State Refinement

The first stage in the TS refinement script is to initiate the running of GRACE. The first line is a comment; all comments in GRACE begin with a "#" . The GRACE code, `grace.tcl` is initiated by the `source` command. The full path to `grace.tcl` must be specified and is done so by using an environment variable, `GRACE_LIB`, usually located in the .cshrc file. On the DEC 8400 at R.A.L., `GRACE_LIB` is set to `/home/columbus_chem/sfc/grace`. Two similar environment variables are also added to the .cshrc file. These define the location of the TCL and CHARMM executables and are `GRACE_CHARMMEXEC` and `TCL_LIBRARY` respectively.

During execution, GRACE makes use of two directories. A "working" directory incorporates all the CHARMM startup files, co-ordinate files and the GRACE script being used. The "scratch" directory contains the hessian and hessian up-date files as well as structures saved in the GRACE format. Hessian and hessian up-date files usually have the extension `.dump`, while structure files are given the `.xyz` extension. The scratch directory can also be defined in the .cshrc file with the `GRACE_SCRATCH` environment variable. The third line of the following code sets the scratch directory to the one defined by `GRACE_SCRATCH`.

```
# Start grace going
source $env(GRACE_LIB)/grace.tcl
set GLOBAL(scratch) $env(GRACE_SCRATCH)
init i
```

The next stage is to use the *CHARMM QM/MM system set up* (appendix 1) to define the system under study. It is important to note that *in GRACE scripts, CHARMM commands begin with a capital letter and are ignored by GRACE while the commands associated with GRACE routines begin with a lowercase letter.* This rule must be adhered to, otherwise the script will fail to execute correctly.

CHARMM QM/MM SYSTEM SET UP

After defining the quantum system in CHARMM (appendix 1), the script prepares to run the GRACE part of the simulation. The first two commands in the following code are CHARMM commands (they begin with a capital letter); they restrict output information to only include essential warnings and data when any CHARMM command is used. The third command, `wait`, instructs GRACE to wait for CHARMM to set-up the QM/MM system.

```
Prnl 0
Wrnl 0

wait
```

The next piece of code defines the `frelax` procedure; this is not executed until the hessian matrix has been generated for the CORE atoms. For it to execute successfully, the CORE must have been already defined. The procedure optimises the ENVIRONMENT after every P-RFO step taken (the latter involving the CORE). The first stage is to obtain the forces and gradients:

```
proc frelax xvec\ gvec {
   upvar $xvec xparam $gvec grad
   global X Y Z WMAI DX DY DZ PARA QMSTART
```

The next stage is to alter the position of the atoms in the CORE; the `silent` option suppresses the data written to the output file.

```
   qmmm silent unmake_pos=xparam
```

CHARMM is now instructed to optimise the ENVIRONMENT around a fixed CORE using the ABNR optimiser. The RMS gradient tolerance is set to 0.005 kcal mol^{-1} $Å^{-1}$ (0.02 kJ mol^{-1} $Å^{-1}$), with an initial `step` size of 0.0001Å; the maximum number of steps permitted, `nstep`, is set to 5000. After the ENVIRONMENT is minimised, the CORE is unconstrained. The comments made by Turner about the `set x` command are left here for completeness.

```
Scal xref stor 1
Scal yref stor 2
Scal zref stor 3
Cons fix  sele  core end
Scal zref reca 3
Scal yref reca 2
Scal xref reca 1
Prnl 5
Mini abnr nstep 5000 tolg 0.005 step 0.0001 nprint 1
Prnl 0
Scal xref stor 1
Scal yref stor 2
Scal zref stor 3
Cons fix  sele none end
Scal zref reca 3
Scal yref reca 2
Scal xref reca 1
set X(1)
# there appears to be an unresolved bug in some cases
# it may have been sorted, but to be sure, hit the
# X cache in ensure upload
set X(1)
```

The energy and gradient of the new system must now be obtained. The energy refers to the total energy, E_{tot} of the QM/MM system while the gradients refer to just the atoms within the CORE.

```
set f [fcharmm xparam grad]
```

The next stage is to write a restart file containing the new atomic positions. The restart file `lats2s1.xyz` is written to the scratch directory as shown by the "<" symbol. The first two lines organise the relevant data into a list that is subsequently sent to the file. All three lines are therefore required for a successful save.

```
xyz>p p=xyz x=X y=Y z=Z
array>list array=xyz list=lxyz
scratch< name=lats2s1.xyz var=lxyz
```

It is often useful to keep track of the geometrical parameters directly associated with the reaction co-ordinate under study. The CHARMM `Quick` command gets this information and GRACE stores it in variable `d1`. This information is then sent to standard output (`flush stdout`); the line "`+++ Distance 2-326 = $d1`" is printed, where `$d1` is the value contained within the `d1` variable. In this case, two geometrical parameters are considered.

```
   Quick 2 326
   set d1 $PARA(DIST)
   puts "+++ Distance 2-326 = $d1"
   Quick 2 38
   set d2 $PARA(DIST)
   puts "+++ Distance 2-38  = $d2"

   flush stdout
```

The `return` command instructs GRACE to call the `frelax` procedure again unless the gradient tolerance has been achieved (or the number of calls has exceeded the maximum number). If any of the exit criteria for `frelax` have been satisfied, GRACE returns to the main part of the script. Finally, "`}`" signals the end of the `frelax` procedure.

```
   return $f
}
```

The code presented from now on is part of the main GRACE script and immediately follows the *CHARMM QM/MM system set up* as far as execution of the script is concerned. The first GRACE command (`qmmm`) defines the atoms to be treated quantum mechanically in the QM/MM system. This is only valid if external quantum mechanical code is being used (such as CADPAC or GAMESS). GRACE labels all CHARMM atoms as MM atoms (even though some are treated with the AM1 Hamiltonian in the program).

```
qmmm \
   fmm=all
```

The atoms to be included in the CORE are now defined (using a CHARMM command). In this case, the solute/substrate (segment identifier `NEU5`) has been

selected to be the CORE. The selection is printed out for each atom, showing the atom symbol (located in the IUPC variable), the residue number (located in the RESI variable) and the residue name (located in the RESN variable). The output of such information is controlled by the for loop in which the variable, i, controls the number of times the loop is repeated. After each execution of the loop code (contained within "{" and "}" brackets), the value of i is increased by one and the information of the next atom is printed.

```
Prnl 5
Defi core sele (segid NEU5) end

Prnl 0
acquire_selection selection=core var=lcore
puts "Selected atoms and types"
foreach i $lcore {
   puts "$IUPC($i) $i $RESI($i) $RESN($i)"
}
flush stdout
```

It is now appropriate to load in a GRACE restart co-ordinate file. This is only required if a previous calculation has been prematurely terminated (in this particular script these lines are treated as comments and are therefore not executed by GRACE). When these lines are uncommented, the restart file soltslsi3.xyz is read from the working directory (shown by the forward arrow ">", which follows the name work). The two lines below read the data contained within the restart file and place it into the variables and arrays used by GRACE. They are therefore required for successful reading of the file.

```
#work> name=soltslsi3.xyz var=lxyz
#list> array array=xyz list=lxyz
#p>xyz p=xyz x=X y=Y z=Z
```

In this particular script, the structure currently being used at this point is the approximate TS structure (determined by adiabatically mapping out a reaction profile). Such a structure is usually read in as a CHARMM .pdb restart file (see appendix 1). The first stage of TS refinement is to make sure the ENVIRONMENT of this structure is minimised while keeping the CORE fixed. The CORE has to be kept fixed because minimisation could result in the system falling into the nearest local minimum. In this script, all variables used in conjunction with the ABNR optimiser

in CHARMM are the same as for the ENVIRONMENT minimisation within the `frelax` procedure.

```
Scal xref stor 1
Scal yref stor 2
Scal zref stor 3
Cons fix  sele core end
Scal zref reca 3
Scal yref reca 2
Scal xref reca 1
Prnl 5
Mini abnr nstep 5000 tolg 0.005 step 0.0001 nprint 1
Prnl 0
Scal xref stor 1
Scal yref stor 2
Scal zref stor 3
Cons fix  sele  none end
Scal zref reca 3
Scal yref reca 2
Scal xref reca 1
set X(1)
```

Very often, this part of the simulation takes a fair amount of CPU time and it is always best to save a restart structure before proceeding. The name of this restart file is `lats2s1_start.xyz` and it is saved to the scratch directory.

```
xyz>p x=X y=Y z=Z p=pos
array>list array=pos list=lpos
scratch< name=lats2s1_start.xyz var=lpos
```

The `qmmm` command declares the positions of the CORE atoms; it is used to change their positions (`make_pos`) in the `frelax` routine. The "\" represents the end-of-the-line and signals to GRACE that a new command is on the next line. If this character is omitted, the script fails to run correctly. If anything is added immediately after the "\" on the same line (including white space characters such as TABS and spaces), the script also fails to run correctly.

```
qmmm \
   mm=core \
   make_pos=pos \
   start
```

The hessian is now obtained for the approximate QM/MM TS structure (in which the ENVIRONMENT has been fully relaxed) using the `gethess` command. The procedure

fcharmm obtains the total energy as well as gradients of the CORE atoms; the positions also refer to just the CORE atoms. The evaluated hessian is then saved to the dump file, lats2s1_start_hess.dump, in the scratch directory. If the generation of the hessian is interrupted, then the RESTART command can be used in place of the dump_name to complete the hessian matrix.

```
gethess \
   func=fcharmm \
   position=pos \
   dump_name=$GLOBAL(scratch)/lats2s1_start_hess.dump \
   hessian=h
```

The matrix is now diagonalised and the resulting eigenvalues are obtained. pmatrix prints the eigenvalues to standard output.

```
la symm matrix=h result=h
la eign matrix=h

pmatrix matrix=values
```

Now the frelax procedure is called to start refining the approximate TS structure. This is done using the optimise command with the func variable set to frelax. The variable nfunc determines the maximum number of calls made to frelax. The P-RFO optimiser follows a chosen eigenvector; it must be locked onto the eigenvector associated with the negative eigenvalue of the TS. The variable, mode, indicates the position in the hessian corresponding to this eigenvalue. For hessian matrices with only one negative eigenvalue, therefore, the mode to be followed will be the first. In addition to these variables, the maximum gradient tolerance is set to 0.3 kcal mol^{-1} $Å^{-1}$ (1.26 kJ mol^{-1} $Å^{-1}$) and the update hessian is saved to the lats2s1_ef.dump file in the scratch directory. Restarting a failed job can be done specifying the RESTART command in place of the dump_name variable; this must be used in conjunction with the RESTART command for the initial hessian. Both files must be located in the scratch directory for a restart to be successful.

```
optimise \
   position=pos \
```

```
func=frelax \
nfunc=100000 \
omin=0.800 \
osmin=0.001 \
rmin=0.90 \
rmax=1.10 \
step=0.001 \
maxstep=0.01 \
mode=1 \
lock \
dump_name=$GLOBAL(scratch)/lats2s1_ef.dump \
method=ef \
hessian=h \
toler=0.3
```

After a successful refinement, the final structure and the initial hessian are saved to the respective `*.xyz` and `*.hes` files in the scratch directory. The structure is then also saved in PDB format using a CHARMM command.

```
xyz>p x=X y=Y z=Z p=final_pos
array>list array=final_pos list=lp
array>list array=h list=lh
la eign matrix=h
array>list array=roots list=lr
scratch< var=lp name=lats2s1_final.xyz
scratch< var=lh name=lats2s1_final.hes

Open unit 67 write form name lats2s1_final.pdb
Write coor pdb unit 67
* Optimized structure
*
Close unit 67
```

The script is then instructed to finish.

```
exit
```

(b). Script for following the Intrinsic Reaction Co-ordinate

Initiate the running of GRACE and apply the *CHARMM QM/MM system set up* procedure.

```
source $env(GRACE_LIB)/grace.tcl
set GLOBAL(scratch) $env(GRACE_SCRATCH)
init i

Bomb -1

                        CHARMM QM/MM SYSTEM SET UP
```

Instruct GRACE to wait until the CHARMM QM/MM set-up has been achieved.

```
Prnl 0
```

```
Wrnl 0

wait
```

The `frelax` procedure is defined in much the same was as for the TS refinement script. It relaxes the ENVIRONMENT after moving the CORE atoms in the IRC calculation.

```
proc frelax xvec\ gvec {

   upvar $xvec xparam $gvec grad

   global X Y Z WMAI DX DY DZ PARA QMSTART

   qmmm silent unmake_pos=xparam

   Scal xref stor 1
   Scal yref stor 2
   Scal zref stor 3
   Cons fix  sele  core end
   Scal zref reca 3
   Scal yref reca 2
   Scal xref reca 1
   Prnl 5
   Mini abnr nstep 1000 tolg 0.005 step 0.0001 nprint 1
   Prnl 0
   Scal xref stor 1
   Scal yref stor 2
   Scal zref stor 3
   Cons fix  sele  none end
   Scal zref reca 3
   Scal yref reca 2
   Scal xref reca 1
   set X(1)
   # there appears to be an unresolved bug in some cases
   # it may have been sorted, but to be sure, hit the
   # X cache in ensure upload
   set X(1)

   set f [fcharmm xparam grad]

   xyz>p p=xyz x=X y=Y z=Z
   array>list array=xyz list=lxyz
   work< name=lats2s1_ircb.xyz var=lxyz

   Quick 2 16
   set d1 $PARA(DIST)
   Quick 2 326
   set d2 $PARA(DIST)
   puts "+++ Distance 2-16  = $d1"
   puts "+++ Distance 2-328 = $d2"

   flush stdout

   return $f
}
```

As with the TS refinement script, the first task is to declare the atoms described by other QM codes. All atoms used in CHARMM are defined as MM atoms (fmm), which in this case is the entire system.

```
qmmm \
   fmm=all
```

Select the atoms in the system to be included in the CORE and print information on the selected atoms to standard output.

```
Prnl 5
Defi core sele (SEGID neu5) end

Prnl 0
acquire_selection selection=core var=lcore
puts "Selected atoms and types"
foreach i $lcore {
   puts "$IUPC($i) $i $RESI($i) $RESN($i)"
}
flush stdout
```

Now read in the refined TS structure file.

```
work> name=lats2s1_final.xyz var=lxyz
list>array array=xyz list=lxyz
p>xyz p=xyz x=X y=Y z=Z
```

Fix the CORE of the refined TS structure and optimise the ENVIRONMENT. The amount of optimisation required is likely to be very small if the refined TS structure is taken from the .xyz file.

```
   Scal xref stor 1
   Scal yref stor 2
   Scal zref stor 3
   Cons fix  sele  core end
   Scal zref reca 3
   Scal yref reca 2
   Scal xref reca 1
   Prnl 5
   Mini abnr nstep 1000 tolg 0.005 step 0.0001 nprint 1
   Prnl 0
   Scal xref stor 1
   Scal yref stor 2
   Scal zref stor 3
   Cons fix  sele  none end
   Scal zref reca 3
   Scal yref reca 2
```

```
   Scal xref reca 1
   set X(1)
```

The qmmm command declares the positions of the CORE atoms; it is used to change their positions (make_pos) in the frelax routine.

```
qmmm \
   mm=core \
   make_pos=pos \
   start
```

The hessian matrix for the refined TS is obtained. Up until this point, no such calculation has been done: the TS refinement calculation generates the hessian for the approximate TS and then updates the hessian after each refinement.

```
gethess \
   func=fcharmm \
   position=pos \
   restart=$GLOBAL(scratch)/lats2s1_ircf_start_hess.dump \
   hessian=h
```

For the IRC calculation, the hessian is mass weighted and the resulting eigenvalues are printed to the output file.

```
acquire_selection selection=core var=sele_list
catch {unset masses}
set c 0
foreach i $sele_list {
   incr c
   set masses($c) $MASS($i)
}
la mass mass=masses matrix=h result=mh
la eign matrix=mh

pmatrix matrix=values

# make velocity the first root
list>array list=roots(1) array=velocity
```

The reaction path is now calculated. Two IRC calculations are done: one leading to reactants, the other leading to the product well. The backward reaction is achieved by scaling the velocities by –1. The first line of the next part of the script achieves this and is not required for the forward IRC calculation. It should be noted that the hessian for the refined TS structure is the same in both IRC calculations and doesn't

need to be re-evaluated. The hessian can be read in by placing RESTART in place of dump_name below the gethess command. The gradient tolerance of 0.1 kcal mol^{-1} $Å^{-1}$ (0.42 kJ mol^{-1} $Å^{-1}$) is rarely achieved. This is because optimising the ENVIRONMENT structure can occasionally result in an increase in the total energy of the system and a premature exit from the IRC calculation. This is why a further optimisation of both the IRC reactant and product structures is carried out (in a separate calculation) using CHARMM.

```
la mbys matrix=velocity scale=-1.0 result=velocity
reaction_path \
   func=frelax \
   nfunc=100000 \
   position=pos \
   velocity=velocity \
   masses=masses \
   print=2 \
   accuracy=15 \
   tstep=1.0 \
   toler=0.1 \
   dump_name=$GLOBAL(scratch)/lats2s1_ircb.dump \
   half=0.0
```

The reactant or product structure is then saved to a .xyz file or a .pdb file.

```
xyz>p x=X y=Y z=Z p=pos
array>list array=pos list=lpos
work< name=lats2s1_ircb.xyz var=lpos

# save end structure in pdb format
Open unit 67 write form name lats2s1_ircb.pdb
Write coor pdb unit 67
* pdb file of backward irc path
* lactone break (react)
Close unit 67
```

Finally, the running of the script is finished.

```
exit
```

(c). Generation of CAMISO input script (camiso.mak)

After following the IRC path and obtaining the reactants and products, it is possible to evaluate the isotope effects associated with the chemical change under study using the program, CAMISO. Before being able to do this, the rotation and translation contamination must be removed from the mass-weighted hessian matrix and the CAMISO input decks for each chemical species must be produced.

The camiso.mak script reads in the co-ordinates of the CORE atoms (in this case ts.pos) and the associated hessian (ts.hes). The atoms contained in the system are listed (in the same order as obtained from the CORE selection) in the symbol.atsy file. The resulting co-ordinate and second derivative data is saved in the CAMISO input deck format (ts.deck). Z is the log file generated as the camiso.mak script is executed. The following command is run from the UNIX command line (without the "#").

```
# tclsh camiso.mak position=ts.pos hessian=ts.hes atsy=symbol.atsy
name=ts.deck > & z &
```

The execution of GRACE is now initiated and a title is output to Z; this includes the names of the .hes, .pos, and .atsy files used to generate the CAMISO input deck. The title provides a useful check to see if the correct files are used.

```
source $env(GRACE_LIB)/grace.tcl
init i n
parse make_camvib "silent $argv" { hessian position atsy  name}
puts "================================================"
puts "             CAMISO INPUT GENERATOR"
puts "              A.J.Turner May 1997"
puts "================================================"
puts ""
puts "==============="
puts "    OPTIONS\n"
puts "hessian  file : $hessian"
puts "position file : $position"
puts "symbol   file : $atsy"
puts "output   file : $name"
puts "\n==============\n"
puts "NATOM = [llength $atsy] \nNHESS = [llength $hessian] \nNCOOR =
[llength $position]"
```

The files are read in as lists from the working directory.

```
work> name=$hessian var=hess
work> name=$position var=lpos
work> name=$atsy var=latsy
```

These lists are now converted to TCL arrays.

```
list>array list=latsy array=symbols
list>array list=hess  array=h
list>array list=lpos array=pos

p>xyz p=pos x=x y=y z=z
```

The hessian must now be scaled into hartree $bohr^{-2}$ for input into CAMISO

```
la mbys  matrix=h scale=[expr 1.0 / 2242.37864102] result=h
```

The rotation and translation components are now removed from the hessian matrix and the resulting eigenvalues are printed out to Z.

```
la prjh p=pos hessian=h result=h
la eign matrix=h nroots=0
pmatrix matrix=values
```

The first part of the CAMISO input deck is to list the atoms and masses with the corresponding Cartesian co-ordinates. For each atom defined in the `.atsy` file, the script places the relevant isotope in the `deck` variable. For example, if the atom in `.atsy` is hydrogen, then "1H" is printed; if it is a carbon atom, then "12C" is printed to `deck`. If there are no atom symbols in the `.atsy` file, then the error message `ERROR ==> Empty atomic symbol entry` will be printed. All the co-ordinates in the CAMISO input deck are reported to seven decimal places. If the atom symbol is not recognised then the error message `ERROR ==> Type '$atsym' not availble to Camvib, use masses instead` is printed (where `$atsym` is the offending atom symbol supplied (It is held in the `atsym` variable)).

```
set deck " 1 1 0 - generated from make_camiso.emp\n"
set c 0
foreach i [array names symbols] {
   incr c
   set atsym $symbols($c)
   case [string tolower $atsym] {
      h  {set s   1H}
      li {set s  6LI}
      be {set s  9BE}
      b  {set s  10B}
      c  {set s  12C}
      n  {set s  14N}
      o  {set s  16O}
      f  {set s  19F}
      si {set s 28SI}
      p  {set s  31P}
      s  {set s  32S}
      cl {set s 37CL}
      ""
      {
         puts "\nCartesian Coordiantes so far:-\n$deck"
         flush stdout
```

```
         error "ERROR ==> Empty atomic symbol entry"
      }

      default
      {
         puts "Cartesian Coordiantes so far:-\ndeck"
         flush stdout
         error "ERROR ==> Type '$atsym' not availble to Camvib, use
masses instead"
      }
   }
   append deck [format "%4s %12.7f%12.7f%12.7f" $s $x($c) $y($c)
$z($c)]\n
}
```

After the co-ordinates, an "*" and new line character are added to the `deck` variable. This symbol separates the co-ordinates and the Cartesian force constants that are subsequently appended to `deck`.

```
append deck "*\n"
set cx 0
set out " "
set nvar [llength $h(1)]
while {$cx < $nvar}  {
   incr cx
   set cy 0
   foreach j $h($cx) {
      incr cy
      if {$cy > $cx} break
      append out [format %20.12e $j]
      case $out in *.*.*.* {
         append deck [set out]\n
         set out " "
      }
   }
}
```

All the information contained within the `deck` variable is the sent to the `ts.deck` file in the working directory.

```
work< var=deck name=$name
```

Appendix 3: Example CAMISO script

The task now is to calculate the isotope effects using CAMISO. The input script of CAMISO first provides the program with information on the type of isotope effect to be calculated (RATE or EQUI) and the number of reactions to be investigated. Each different isotopic substitution is regarded as a single reaction. To evaluate four isotopic substitutions, five reactions must be carried out (this includes the unsubstituted reaction).

```
ISO
RATE 5
```

After the title line, the absolute temperature is supplied (in this case 310.15K). Below that, the barrier height is supplied for the process under study; it is used to evaluate any tunnelling corrections needed for the system. The first number on this line is a flag that can assume a value of zero or one; it determines if the barrier height is included into the calculation. The next two lines give the species studied and the relative stoichiometry. In this case, one reactant molecule leads to one TS molecule.

```
ISOTOPIC RATE CONSTANTS FOR TS
 1     310.15
 0      6.50
 1 REACT
 1 TS
```

The co-ordinates and force constants are now supplied for the reactant. The "1 1 0" line states the rotational symmetry number and the electronic degeneracy (or multiplicity) of the reactant molecule; the zero is a flag and is set to one if vibrational frequencies are supplied as input data. The rotational symmetry number represents the number of ways a molecule can be rotated into an identical orientation. For the ES complexes, both the symmetry number and multiplicity is set to one. There are no blank lines between the "1 1 0" or the previous line (i.e. the 1 TS line). Similarly there is no newline character between the "1 1 0" and the reactant input deck. The braces only indicate the start and end of the reactant input deck and do not occur in the CAMISO script itself. The "0 2 1 2" line informs CAMISO of respectively, the number of imaginary and the number of near-zero frequencies in

the system. In this case, there are no imaginary frequencies and two near-zero frequencies in the reactant (these being the first and second reported frequencies in the system).

```
1 1 0
{
      Reactant CAMISO input deck
}
0 2 1 2
```

Next in the CAMISO script is the TS structure information; it immediately follows the "0 2 1 2" line of the reactants. This time, however, there is one imaginary frequency and two near-zero frequencies associated with the TS. The latter are reported as the second and third frequency of the system.

```
1 1 0
{
      TS CAMISO input deck
}
1 2 2 3
```

After reading in the TS information, the other reactions are defined with the stoichiometry of the system given first. All the atomic masses of the atoms in the system are then supplied with a maximum of 16 atoms declared on one line before moving onto the next. The "." in the following part of the script represents a "space"; there should also be no white space characters inbetween the reactions. Any deviation from this format will result in the script not running successfully. The isotopically substituted species for each reaction is shown in bold.

Reaction 2:

```
1 REAC
1 TS
1 1 0  PRO-R EFFECT
.12C..12C..16O..12C..12C..12C..16O..12C..16O...1H...1H...1H...1H...1H...2H..16O
.12C..16O...1H...1H..12C..12C..16O..16O...1H...1H...1H...1H...1H..14N..12C..12C
..1H...1H...1H..16O...1H..16O..12C..12C..12C..12C..14N..12C..12C...1H...1H..16O
.16O...1H...1H..12C...1H...1H..12C..16O..16O...1H
0   2   1   2
1 1 0
.12C..12C..16O..12C..12C..12C..16O..12C..16O...1H...1H...1H...1H...1H...2H..16O
.12C..16O...1H...1H..12C..12C..16O..16O...1H...1H...1H...1H...1H..14N..12C..12C
..1H...1H...1H..16O...1H..16O..12C..12C..12C..12C..14N..12C..12C...1H...1H..16O
.16O...1H...1H..12C...1H...1H..12C..16O..16O...1H
1   2   2   3
```

Reaction 3:

```
1 REAC
1 TS
1 1 0  PRO-S EFFECT
.12C..12C..16O..12C..12C..12C..16O..12C..16O...1H...1H...1H...1H...2H...1H..16O
.12C..16O...1H...1H..12C..12C..16O..16O...1H...1H...1H...1H...1H..14N..12C..12C
..1H...1H...1H..16O...1H..16O..12C..12C..12C..12C..14N..12C..12C...1H...1H..16O
.16O...1H...1H..12C...1H...1H..12C..16O..16O...1H
0   2   1   2
1 1 0
.12C..12C..16O..12C..12C..12C..16O..12C..16O...1H...1H...1H...1H...2H...1H..16O
.12C..16O...1H...1H..12C..12C..16O..16O...1H...1H...1H...1H...1H..14N..12C..12C
..1H...1H...1H..16O...1H..16O..12C..12C..12C..12C..14N..12C..12C...1H...1H..16O
.16O...1H...1H..12C...1H...1H..12C..16O..16O...1H
1   2   2   3
```

Reaction 4:

```
1 REAC
1 TS
1 1 0  COMBINED PRO-R and PRO-S EFFECT
.12C..12C..16O..12C..12C..12C..16O..12C..16O...1H...1H...1H...1H...2H...2H..16O
.12C..16O...1H...1H..12C..12C..16O..16O...1H...1H...1H...1H...1H..14N..12C..12C
..1H...1H...1H..16O...1H..16O..12C..12C..12C..12C..14N..12C..12C...1H...1H..16O
.16O...1H...1H..12C...1H.. 1H..12C..16O..16O...1H
0   2   1   2
1 1 0
.12C..12C..16O..12C..12C..12C..16O..12C..16O...1H...1H...1H...1H...2H...2H..16O
.12C..16O...1H...1H..12C..12C..16O..16O...1H...1H...1H...1H...1H..14N..12C..12C
..1H...1H...1H..16O...1H..16O..12C..12C..12C..12C..14N..12C..12C...1H...1H..16O
.16O...1H...1H..12C...1H.. 1H..12C..16O..16O...1H
1   2   2   3
```

Reaction 5:

```
1 REAC
1 TS
1 1 0  LEAVING GROUP O
.12C..12C..16O..12C..12C..12C..16O..12C..16O...1H...1H...1H...1H...1H...1H..16O
.12C..16O...1H...1H..12C..12C..16O..16O...1H...1H...1H...1H...1H..14N..12C..12C
..1H...1H...1H..16O...1H..18O..12C..12C..12C..12C..14N..12C..12C...1H...1H..16O
.16O...1H...1H..12C...1H...1H..12C..16O..16O...1H
0   2   1   2
1 1 0
.12C..12C..16O..12C..12C..12C..16O..12C..16O...1H...1H...1H...1H...1H...1H..16O
.12C..16O...1H...1H..12C..12C..16O..16O...1H...1H...1H...1H...1H..14N..12C..12C
..1H...1H...1H..16O...1H..18O..12C..12C..12C..12C..14N..12C..12C...1H...1H..16O
.16O...1H...1H..12C...1H...1H..12C..16O..16O...1H
1   2   2   3
```

Finally, the `STOP` command is issued to terminate the running of the script.

```
STOP
```